APRIL
1960
VOLUME 24
NUMBER 2

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

EDWARD S. BORDIN, Editor
University of Michigan

BOYD MCCANDLESS, Associate Editor
University of Iowa

Advisory Editors: John W. Atkinson, *University of Michigan.* Frank X. Barron, *University of California, Berkeley.* John M. Butler, *University of Chicago.* Allen T. Dittman, *National Institute of Mental Health.* Seymour Fisher, *Baylor University.* Harrison G. Gough, *University of California, Berkeley.* Robert B. Malmo, *McGill University.* Paul E. Meehl, *University of Minnesota.* Daniel R. Miller, *University of Michigan.* Ruth L. Munroe, *New York, N. Y.* Saul Rosenzweig, *Washington University.* Julian B. Rotter, *Ohio State University.* Nevitt Sanford, *University of California, Berkeley.* Seymour B. Sarason, *Yale University.* Robert R. Sears, *Stanford University.* Edwin S. Shneidman, *Veterans Administration, Los Angeles, California.* Read D. Tuddenham, *University of California, Berkeley.*

The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

Manuscripts submitted for *regular* or *early* publication, Brief Reports, and tests for review should be sent to the Editor, Edward S. Bordin, Department of Psychology, University of Michigan, Ann Arbor, Michigan. All manuscripts should be submitted *in duplicate* to facilitate referral to the Advisory Editors.

Style of Manuscripts. Manuscripts must be prepared in the style described in the *Publication Manual of the American Psychological Association (1957 Revision)*, obtainable for $1.00 from the Washington office of the APA.

Early Publication. Authors may secure the early publication of an acceptable article in the first issue to go to press after its acceptance by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. Information about charges for early publication may be obtained from the Managing Editor.

American Documentation Institute. To save printing costs, materials supplementary to articles may be deposited with the ADI. To obtain copies, order by Document Number as cited in the reference footnote, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. The indicated fee must be remitted in advance. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Reprints. Authors receive fifty reprints gratis. Additional quantities may be ordered when returning proofs.

Arthur C. Hoffman
Managing Editor

Helen Orr
Promotion Manager

Virginia Daly
Editorial Assistant

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must reach the subscription office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

The *Journal of Consulting Psychology* is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 576 pages. The subscription per year is $8.00, foreign $8.50, single number $1.50.

Published by the American Psychological Association, Inc.
at Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth St. N.W., Washington, D.C.

Second class postage paid at Lancaster, Pa.

© 1960 by the American Psychological Association, Inc.

JOURNAL OF CONSULTING PSYCHOLOGY

April 1960 Vol. 24, No. 2

Contents

ARTICLES

Intelligence: A Concept in Need of Re-examination: *Shephard Liverant* - - - - - 101

Personality Differences and Level of Aspiration: *June Elizabeth Chance* - - - - - 111

Manifestations of Hostility in Neurodermatitis: *M. Michael Klaber* - - - - - - 116

A Comparison of Changes in Psychoneurotic Patients During Matched Periods of Therapy and No Therapy: *Rosalind Dymond Cartwright and John L. Vogel* - - - 121

Toward the Elimination of the Concept of Normality: *David Freides* - - - - - - 128

Age Differences in the Interscale Factor Structure of the Guilford-Zimmerman Temperament Survey: *A. W. Bendig* - - - - - - - - - - - - - - - - - - - 134

Size-Distance Judgment in Organic Mental Defectives: *Noël Jenkin and Sally A. Morse* 139

Impulse Repression and Emotional Adjustment: *Harry Grater* - - - - - - - - 144

Critical Flicker Frequency and the Spiral Aftereffect with Process and Reactive Schizophrenics: *Joseph M. McDonough* - - - - - - - - - - - - - - - - - - 150

An Attempt to Employ a Sentence Completion Test for the Detection of Psychiatric Patients with Suicidal Ideas: *Herman Y. Efron* - - - - - - - - - - - - - 156

Factor Analyses of "Anxiety" and "Neuroticism" Inventories: *A. W. Bendig* - - - 161

Perception of Maternal Child Rearing Attitudes in Schizophrenics: *Alfred B. Heilbrun, Jr.* - 169

Ambiguity and Discriminating Power in Personality Inventories: *William E. Broen, Jr.* 174

Implicit Personality Theories of Clinicians as Defined by Semantic Structures: *Maurice Korman* - 180

WISC IQs for the Mentally Retarded: *Donald P. Ogdon* - - - - - - - - - - - 187

Incidence of Bender-Gestalt Figure Rotations: *Richard M. Griffith and Vivian H. Taylor* - 189

BRIEF REPORTS

A Factor Analytic Study of the Self-Concept: *Philip A. Smith* - - - - - - - - 191

The Effects of Chlorpromazine on Psychological Test Scores: *Abe J. Judson and Barbara W. MacCasland* - 192

Simultaneous Recording of Bales and Chapple Interaction Measures During Initial Psychiatric Interviews: *A. Paul Hare, Nancy Waxler, George Saslow, and Joseph D. Matarazzo* - 193

Social-Psychological Factors and the Psychiatric Complaints of Disturbed Children: *Raymond G. Hunt, Jack L. Roach, and Orville Gurrslin* - - - - - - - - - 194

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.

2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.

3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.

4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual.* Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota:

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. ——, remitting $—— for microfilm or $—— for photocopies.

Extended report. Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (*a*) The extended report should be typed with single spacing for economy in duplication. (*b*) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (*c*) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual.*

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

INTELLIGENCE: A CONCEPT IN NEED OF RE-EXAMINATION[1]

SHEPHARD LIVERANT
Ohio State University

The importance of the concept of intelligence in understanding and predicting human behavior appears so obvious that the value of this term in a science of behavior has seldom been questioned by psychologists. The inadequacies inherent in present conceptions of intelligence have been made explicit by a number of authors (Anastasi, 1958a; Coombs, 1952; Fromm & Hartman, 1955; McClelland, 1951), but the reader is left with the impression that the concept itself is essentially a useful and necessary one. By leaving open the possibility that the concept may be unnecessary, and even an obstacle, in explaining behavior, perhaps we can arrive at a better solution to the problems now subsumed under the rubric of intelligence.[2]

This paper explores a number of problems over which the intelligence concept has stumbled, such as: establishing an unequivocal definition, the contribution of heredity and environment, measuring an innate capacity, situational predictions, the constancy of the IQ, intelligence as a general factor, the role of culture in defining intelligent behavior, and the function of nonintellective variables. The objective of such an exploration is to re-evaluate these persistent problems by analyzing the logical and methodological adequacy of the meanings carried by intelligence.

Present uses of the term intelligence for individual prediction and the difficulties inherent in such usage are best represented in the clinical setting. Consequently a brief look at the clinical approach to intelligence will serve as a point of departure for the subsequent analysis.

Clinical Approach

Clinical case studies have produced a greatly increased awareness of the failure of an IQ (or any other behavioral appraisal) to reflect accurately a constant ability to perform at a certain level in all situations. This awareness has led to the clinical practice of considering the IQ as an index of present intellectual functioning rather than intellectual potential or capacity. However, the belief in the existence of a "true" IQ or intelligence persists, and attempts to discover it in each individual continue.

Largely on the basis of scatter or pattern analysis of intelligence tests, statements are made concerning the size of the discrepancy between what is now present in the way of intellectual functioning and some assumed absolute capacity. Were uneven test profiles produced only by deviant individuals such interpretations might be meaningful, but subtest scatter is the rule rather than the exception within any population in spite of efforts to build homogeneous measures (Shofield, 1952). As will be argued later, it is probably more fruitful and certainly more empiri-

[1] This paper was written in part at the Interdisciplinary Program in the Behavioral Sciences at the University of New Mexico, Summer 1958. The conference was sponsored by the Behavioral Sciences Division, Air Force Office of Scientific Research under Contract AF 49 (638)-33. Their support is gratefully acknowledged.

The author also wishes to express his appreciation to Julian B. Rotter, Richard Jessor, and Edwin Barker for their invaluable suggestions and criticisms.

[2] In order to forestall misunderstanding, it is emphasized at the outset that the term intelligence is being questioned in its explanatory usage, not necessarily as a descriptive concept. In other words, intelligence is used in two ways; as a descriptive term (e.g., A's behavior is more intelligent than B's) and as an explanation (e.g., A solved more problems than B *because* A is more intelligent). It is the validity of the latter use which is being questioned.

cally justifiable to regard uneven test profiles as the result of conditions (not necessarily abnormal ones) which have differentially affected the acquisition of certain skills and/or the manifestation of these skills in certain classes of situations. Then the task becomes one of determining what these conditions are and how they operate, rather than ascribing subtest patterns to various types of disease entities which affect another entity (intelligence).

In general it appears that the host of evidence gathered over the last twenty or more years challenging the general and innate nature of intelligence has been subsumed under an interference hypothesis. Consequently, intelligence can be and still is viewed as a general and innate source of variance, but it is increasingly recognized that in order to get at it from behavioral observation one must strip away an ever increasing number of complex variables which interfere with its effects on behavior.

This point of view is illustrated by the work of the Chicago group (Eells, Davis, Havighurst, Herrick, & Tyler, 1951) who, after demonstrating the effects of culture on test performance, conclude that the answer to the problem is to develop "culture-free" intelligence tests. By contrast, Anastasi (1950) uses the same kind of evidence to question the validity of any test as a measure of an abstract conception of intelligence without regard to validation criteria. As Rotter (1958, p. 19) points out, "It is somewhat surprising that the emphasis on culture and training did not result in a discarding of the notion of over-all intelligence and the IQ in favor of a clear understanding of the value of varied tests in assessing what has been learned and what each test can predict. Rather the early trend to entitize and reify the IQ continues." (Quoted by permission of McGraw-Hill.)

The question being raised here is what if anything will be left as a contributor to understanding behavior on a psychological level once the stripping away process is complete. If more and more of the variance which accounts for individual differences in intellectual performance can be attributed to other variables, what is intelligence?

The Problem of Definition

Spiker and McCandless (1954) attempt to resolve the definitional problem by an appeal to early operationalism. In effect they conclude that intelligence is what different intelligence tests measure in different situations with different groups and the relationship of these measures to other measures. Their solution reduces intelligence to the status of an intervening variable or pure disposition term as described by Carnap (1956). There is nothing wrong with such a definition of intelligence except that most people are not willing to accept it. Intelligence means more to most of us than a score on an intelligence test even though the "more" hasn't been made explicit (i.e., we use intelligence as a hypothetical construct whose validity depends on the verification or refutation of predictions derived from the theory which contains it).

More recent thinking on the subject suggests that the ultraempirical approach is not a satisfactory solution to the problem of definition. Several authors (Cronbach & Meehl, 1955; Jessor & Hammond, 1957; Liverant, 1958) have argued that the construction, interpretation and validity criteria of any test depend on the nomological network (meanings) surrounding the implicit or explicit construct(s) which the test is designed to measure. Intelligence like any other construct takes its meaning from the assumed antecedent conditions and subsequent responses which it mediates. The controversies over nature-nurture on the antecedent side and general versus specific factors on the response side of the relationship characterize the difficulties in arriving at an accepted definition of intelligence. Thus, the next step is to examine first the nature of the antecedent conditions and then the subsequent responses subsumed by intelligence.

Antecedent Side of the Problem

Heredity and Environment

As now evolved, the antecedents of intelligence do not refer to a set of conditions (either inherited or environmental) but rather to a conceptual model involving postulated genetic operations. Although stated in many different forms, genetic intelligence seems to

embody the idea that each organism, from the moment of conception throughout the life span, has inherent within it an unmodifiable growth process determined by gene composition which if given the proper environmental stimulation at the right time will result in maximum problem solving performance for that organism under suitable conditions of motivation, health, etc. Improper stimulation (degree and kind unspecified) may inhibit certain manifestations of this innate property, but given an adequate restoration of environmental conditions (provided structural damage has not occurred) it will proceed upon its genetically determined course. However, no amount of environmental manipulation will serve to modify the limits of intellectual performance prescribed by the genetic constituents. It is this last proposition which lends an air of crucial significance to this formulation, since knowledge of these innately fixed limits (intelligence) will permit us to predict maximum attainment under optimal conditions—predictions which are invaluable in terms of educational and vocational planning.

There are several implications of genetic intelligence which need clarification, if not complete refutation, in light of recent conceptual advances in both genetics and psychology. Paramount is the notion of intellectual limits set by heredity. As usually interpreted this idea means that beyond a certain point (usually unspecified) any and all environmental conditions cease to make a difference as far as increments in problem solving proficiency are concerned. But the conclusions of modern day geneticists (e.g., Huxley, 1949; Stern, 1952) reveal a basic fallacy inherent in this line of reasoning. Geneticists postulate that the appearance of any characteristic (behavioral or otherwise) depends upon the *interaction* of a *specific* genetic structure operating within a *specific* environment. In this sense genes set the limits within any given environment but these limits may vary as the environment varies. Without in any way arguing against the Wiesman-Morgan theory of the immutability of the germ plasm, it is no more logical to attribute a limiting role to heredity than it is to assign the same role to the environment. Perhaps the irrevocability of genetic determinants as compared to situational ones has contributed to this mode of thought, but it is one which contradicts the postulated interaction effect.

The implication of a true interaction is not that only within a limited range (set by one of the variables) does the effect of the other make a difference, but that they both operate to produce their effects over any conceivable range. Now it does appear that heredity can influence the variety of conditions which will or will not have an appreciable effect on a given characteristic. For example, it can be said that species characteristics are determined by heredity in the sense that the occurrence or nonoccurrence of an extremely wide range of external conditions does not make a difference in the morphology necessary to identify a given species. Both theoretically and empirically, however, a set of conditions could be and, in fact, have been produced which radically alter species characteristics. Feeding thyroid abstract to the young axolotl to produce a creature adapted for the land rather than the usual aquatic life is a case in point (Jennings, 1930). On the other hand, intraspecies behavior is apparently affected by a much greater host of nongenetic variables. Furthermore, the importance of situational variations in determining individual differences seems to increase as one ascends the phylogenetic scale. Beach's (1954) findings regarding comparative variations in sexual behavior emphasize the ever increasing role of situational factors in determining the sexual patterns displayed by increasingly complex organisms. It is likely that the same relationship is even more predominant in the case of complex problem solving behavior. Since the evidence points to such a great number of circumstances which can affect intellectual performance it becomes conceptually difficult to speak of heredity as the determiner of intellectual limits, particularly in the case of human beings.

From this line of reasoning it follows that although no practical way may be known to affect appreciable positive changes in an individual's performance level, it cannot be assumed that such changes are precluded by genetically determined limitations. On the contrary, the implication of genetic theoriz-

ing in this respect is that changes in any characteristic would result from sufficient environmental changes. After producing previously unknown patterns of performance in pigeons (i.e., bowling behavior), Skinner (1958, p. 96) concludes that, "It is dangerous to assert that an organism of a given species or age *cannot* solve a given problem." The implications of this reasoning in the field of mental deficiency are apparent.

A further failure to recognize the significance of the continuous interaction between genes and environment in producing behavior is indicated by the one-way effect of interaction implied by the intelligence model. The implication is that training can only have an effect after neural maturation paves the way and that it is by controlling the maturational rate that heredity plays its vital role in the development of intelligent behavior. The "Johnny and Jimmy" type of studies (McGraw, 1935) are cited as evidence in this connection. But this approach neglects the reciprocal relationship between physiological maturation and behavior. The "learning how to learn" phenomena demonstrated by Harlow (1949) strongly suggests that the rate of acquisition of problem solving ability is directly dependent on neural changes which occur as a function of previous problem solving performance. In other words, previous experience as well as genes "shapes up" the organism's maturational readiness for new training procedures.

Acceptance of the interaction thesis concerning heredity and environment invalidates the concept of innate capacity unless the term includes the operations of an environment. Consequently, the prediction of behavior based on genetic antecedents involves the problem of specifying situational conditions, a problem with which the present definition of intelligence as an inherited capacity has not dealt. There are two kinds of situational conditions which need to be specified: first, the conditions of development, growth and learning, and second, the criterion conditions (i.e., the actual performance situations). The emphasis in this section is on the former conditions. The criterion conditions will be discussed more fully in the subsequent response section.

Environmental Specificity

Perhaps the most pronounced conceptual weakness of the present model lies in its inability to cope with environmental specificity. The position that individual differences in problem solving are a function of heredity given a relatively constant "demographic" environment (e.g., similar cultural opportunities or similar family life) grossly oversimplifies the specific and still unknown effects of the "psychological" environment on the hereditary potential.

Recent research concerned with the effects of early experience upon later behavior has brought into sharp focus the preeminent role of experiential factors in influencing problem solving processes. For example, evidence has been gathered that rats (Beach & Jaynes, 1954), dogs (Clarke, Heron, Fetherstonhaugh, Forgays, & Hebb, 1951), and humans (Stone, 1954) reared in impoverished perceptual and social environments show "irreversible" problem solving impairment in adulthood. Extensive clinical case material now supported by still tentative research findings (Axline, 1949) indicates that much more is involved in a child's inability to profit from the most "well conducted" training procedures than lack of the proper genetic constituents. The crucial role of interpersonal relationships in the acquisition of problem solving behavior is becoming increasingly evident.

In order to avoid confusion, it is important to recognize that no amount of empirical evidence relating intelligent behavior or the lack of it to environmental events can directly invalidate the model in question. As now conceptualized the position concerning genetic determinants is that they exist and somehow make a difference, a position which it would be absurd to deny. The most that data can accomplish is to question the usefulness of the genetic intelligence scheme and indicate lines along which a potentially more productive one may develop. The network of studies beginning to elucidate the complex set of interrelated past and present situational factors involved in determining the adequacy of intelligent behavior tends to place a heredity explanation of problem solving ability in the position of an appeal to ignorance. In

the present stage of conceptualization in this area, it is the absence of knowledge concerning situational determinants which brings forth genetic ones.

The Constancy of the IQ

In the final analysis the value of any scientific approach must be judged in terms of the number of valid hypotheses which it generates. The questionable validity of one of the two major hypotheses stipulated by the genetic theory of intelligence (i.e., intelligence as a general factor) will be discussed in the subsequent response section. The inability to validate or refute the other major hypothesis concerning the constancy of the IQ represents still another instance of the essential untestability of the genetic model.

It may well be that given a theoretically constant environment individuals would maintain the same relative position in problem solving adequacy over any given time period as a function of genetic differences. However, the indications that the *inconstancy* as well as the *constancy* of the IQ can be better explained by alternative hypotheses (other than genetic ones) seriously challenge the existing model. Disregarding individual differences in patterns of IQ change, the empirically observed "constancy" of intellectual status at the upper childhood age ranges has been accounted for on the basis of an "overlap" hypothesis. Anderson (1940) among others has proposed that the similarity in IQ from one age level to the next depends on the proportion of similar test items included in the total score at various age levels.

A more crucial test of the constancy hypothesis is provided by the longitudinal studies (Bayley, 1955) explicitly considering individual differences. These conclusively demonstrate the lack of IQ constancy over any childhood period. Sontag, Baker, and Nelson (1958), in probably the most thorough and well designed study of its kind, found that 62% of their children changed more than 15 IQ points in either direction sometime during the developmental period from the age of three through ten years. Most significant is the finding that the direction and degree of change in rate of "mental growth" are related to personality dimensions (primarily a need for achievement).

Reductionism

It can and has been argued that environmental explanations in and of themselves must always remain incomplete since genetic differences undoubtedly do play a role. The demand for the incorporation of genetic operations in behavioral explanatory conventions brings us to the problem of reductionism, i.e., the question of the necessity and/or feasibility of seeking explanations of the phenomena described in the language of one discipline by reference to the descriptive language of another. Reductionism, as far as psychology is concerned, usually refers to attempts to explain molar behavior in terms of physiological (i.e., neurological) events. Since genetics is still lower in the hierarchy, in the sense that its effects on behavior must be mediated through physiological processes, the criticisms cited in reference to physiological reductionism should be even more pertinent in the case of genetic reductionism. Jessor (1958) cogently argues against the logical possibility of reducing constructs designed to abstract organism–environment interactions to physiological operations, since the latter lack terms to describe the behavioral environment. This argument seems to be particularly applicable to any attempts to seek the causal antecedents of problem solving behavior (certainly a complex organism–environment interaction) in genetic laws (certainly lacking terms to describe the behavioral environment).

Whatever the eventual relationship between genetic determinants and complex behavior proves to be, it is obvious that this relationship is a very distant one.[3] Bridging this gap necessitates taking into account a number of intervening processes at different levels of organization. And the simpler or more molecular the process, the closer is the relationship to the genes. Thus, initial sensitivity to stimuli may be more genetically determined than differential responding to these stimuli. Since these processes preceding complex performance are continuously changing as a function

[3] Anastasi (1958b) elaborates this same point in much more detailed and analytic fashion in a recent article.

of innumerable environmental variations in large part brought about by the behaving organism, the more complicated the behavior, the greater is the potential error in prediction based on genetic antecedents.

In this sense, it is not at all surprising to find a lack of correlation (Bayley, 1955) between infant scales which of necessity sample lower (more molecular) levels of behavior and intelligence tests dependent on higher order processes. But the major conclusion to be drawn from our attempts to develop infant scales and culture free tests is that the more we narrow our performance sample to eliminate or minimize the effects of environmental variables, the less likely we are to arrive at valid measures of intelligent behavior as we now conceptualize it.

The enormous difficulties involved in establishing predictive relationships between genes and complex behavior raise the question of the usefulness of any single construct which attempts to mediate such a relationship. Fuller's (1954, p. 16) statement is relevant in this connection, "On *a priori* grounds a biologist would suspect that intelligence would not be a very good subject for genetic analysis, since it is a complex character depending upon the coordination of many separate processes."

Subsequent Response Side of the Problem

Analogous to the previous discussion of its antecedents, intelligence does not refer to a well defined set of subsequent responses, but to an elusive set of notions concerning problem solving. Postulating intelligence as the general something (force, energy, capacity, potential) which undergirds the abilities necessary to solve problems can, of course, neither be directly validated nor invalidated. However, by examining the consequences which have followed from this formulation, one may contrast its usefulness to that of other frames of reference not including intelligence in the above sense as the basis for understanding problem solving.

Intelligence as a General Factor

Analysis of intelligence as a psychological construct raises the problem of pointing to the behavioral referents which are placed in the class, intelligence. The same analysis also requires determination of the grounds upon which certain aspects of behavior are classified. For, depending upon the basis of classification, the invented construct may fulfill its goal in expediting the understanding of behavior or as suggested by Peak (1953, p. 248) it may prove, ". . . artifactual and without systematic significance." Unfortunately it is not clear which are the behavioral referents for intelligence or which principles are involved in isolating the intelligence concept.

Intelligence as a prescientific concept (i.e., its meaning in everyday language) apparently refers to an unspecified but certainly very wide range of problem solving abilities. It is important to note that this usage does not refer directly to the behavior involved in solving problems but to the adequacy of the results of the behavior. In this sense the immediately distinguishable referents for intelligence are not *behaviors* but *problems*. The importance of all this lies in the recognition that selecting a sample of problems from an unknown universe of problems on some historical, cultural and/or common sense basis to measure intelligence (a procedure traditionally followed in intelligence test construction) may involve a number of behaviors which are not functionally related. However, the construct validity of intelligence as referring to all intellectual problems (i.e., a general factor) depends upon the functional relatedness of all or at least most of the behaviors involved in problem solving. That is, in terms of one set of operations, the describable events within the class, intelligence, must correlate as a function of a common source of variance.

Although factor analysis has repeatedly demonstrated the unrelatedness of a striking number of behaviors traditionally grouped under intelligence, there still persists the hypothesis of a general ability which now, as the result of training, interest, motivation, etc., has become differentiated into the various abilities demonstrated by correlational methods. The persistence of a general factor is based on the low but positive correlations often found between the factors which emerge from the test matrix (i.e., a second-order fac-

tor). However, even discounting similarities inherent in the content of the tests usually placed in the matrix, it cannot be assumed a priori that these correlations confirm the existence of general intelligence as usually understood. The communality which is found may, for example, be the result of the testing situation rather than any inherent property within the individual which transcends all situations. A series of studies conducted by Sarason and his coworkers (Gordon & Sarason, 1955; Mandler & Sarason, 1952; Mandler & Sarason, 1953; Sarason & Mandler, 1952) indicates that variables (test taking attitudes or test taking anxiety) operating as a function of the interpersonal competitive situation represented by the test may contribute significantly to test performance. It seems reasonable to hypothesize that these situational variables may to a large extent account for a general factor, and surely, this is not what is meant by intelligence. By systematically varying the conditions under which the same tests are given to the same individuals, it may be possible to demonstrate concomitant changes in the measures as a function of these situational variables.

Regardless of the accuracy of the hypothesis concerning the role of situational factors in accounting for G, the criticism of the existence of a general factor based on factor analytic methods stands. Guilford (1956, p. 290) in an extensive review of factor analytic studies of intellectual tests concludes that, "The methods of multiple factor analysis . . . do not find a general psychological factor at the first order level and they find no second-order factor that can properly lay claim to the title of intelligence."

It is becoming increasingly clear that constructs which explicitly attempt to abstract functionally related behaviors in a situational class are potentially more useful than constructs which do not. For example, Liverant (1958) found increased predictability by dividing the need for recognition into the need for recognition in academic situations, and the need for recognition in social situations. It is reasonable to assume that greatly lowering the level of generality of intelligence by subcategorizing it in terms of specific classes of situations will significantly increase accuracy of prediction. This likelihood has long been recognized within such categorization as social intelligence, academic intelligence, mechanical intelligence, etc. But somehow, we persistently return to a general intelligence, perhaps, because the meaning of the term proclaims the existence of one.

Cultural Definition

Whatever the most useful classification of intellectual behaviors proves to be, the very generality of intelligence makes it impossible actually to sample a unidimensional population of behaviors subsumed by intelligence since in one way or another most, if not all, of behavior sooner or later becomes involved in problem solving. Even scratching one's epidermis may be viewed as solving the problem of itching. However, such behavior tends not to be included in the typical intelligence test. In fact, the criterion of what is or what is not intelligent behavior involves a cultural value judgment. This makes for a great deal of difficulty in arriving at a consistent sample since by definition values are different in different cultures and tend to change with time in a given culture.

Rapaport (1953) suggests that in the case of terms having value connotations it is meaningless to ask what is X. Rather our problem becomes one of finding commonly accepted criteria for X. According to this point of view, rather than asking what is intelligence, we should be seeking established criteria for intelligent behavior which will vary from one class of situations to another. In this sense intelligence is divorced from its explanatory significance and serves a descriptive function, a state of affairs which may prove very desirable.

The Role of Nonintellective Variables

Reference to the situational context in which intellectual behaviors operate brings into focus the role of variables other than the intellectual behaviors per se in the problem solving process. While it is generally conceded that nonintellective factors are important in all actual measures of problem solving ability, within the intelligence model no clear conceptualization of what these variables are and how they function has yet been made. As

testers, we somehow assume that by holding nonintellective variables constant (vaguely referred to as establishing rapport or providing the proper conditions) during the testing period, an index of an individual's fundamental capacity to solve problems can be obtained apart from these nonintellective factors. But the manifestation of any problem solving ability in or out of the testing situation must be in accord with certain laws of behavior which include the operation of these so-called nonintellective variables.

An example using the terms of Rotter's (1954) Social Learning Theory may help clarify this latter point. In social learning theory the probability of the occurrence of any behavior depends upon the individual's expectation that using this behavior in the present situation will result in a certain kind of reward and the value of that reward for him. Thus, an individual may not manifest a certain implicit or explicit skill (e.g., memory, judgment, reasoning, etc.) necessary to solve a given problem for either of two reasons: one, the level of skill required is not within his repertoire; or, two, the right combination of expectancy and reward value necessary to elicit that behavior in that situation is not operating.

Neither intelligence tests nor the intelligence model provides an objective means of making a decision regarding which of the two reasons is operating in the case of failure to solve problems. Wechsler (1958) attempts to circumvent this problem by including the operation of nonintellective variables within his practical definition of intelligence. In practice, however, it is the inability to identify and formulate the operation of these variables which results in the lower predictive validity of intelligence tests, since it is these very variables which change as the situation changes. As a consequence of these situationally modifiable variables we have the fairly common case of the child who functions so well in school but behaves so "stupidly" with his peers or, what is even more puzzling within the confines of intelligence, does admirably on an intelligence test and abominably in school. It is only within a theory of problem solving which attempts an explicit formulation of all the variables necessary to predict behavior that these apparent contradictions become understandable.

Mental Deficiency

An adequate discussion of mental deficiency in light of the present suggestions concerning intelligence is beyond the scope of this presentation. It is mentioned at this point, however, because it is in relation to the intellectually handicapped that the discarding of intelligence as an explanatory construct may have its most far reaching effects. Fortunately, a number of prominent students (Hutt & Gibby, 1958; McCandless, 1952; Sarason, 1959) of this problem have anticipated and demonstrated the detrimental fallacy of regarding certain kinds of problem solving inadequacies as largely a matter of an inherited lack of capacity to perform any and all tasks at a certain level of difficulty. However, continued use of such terms as limited intelligence, inferior intellectual capacity, low IQ, feebleminded, mental defective, etc. (all referring to lack of intelligence as *an explanation* for inadequate performance) hinders the understanding and the development of the point of view being expressed by these authors because of the surplus meanings inherent in the term intelligence. What is necessary from our point of view is recognition of the logical basis for discarding a contradictory and outmoded terminology which persists through tradition and not through theoretical and/or empirical justification.

An Approach to Change

Inherent in the reasons for relegating intelligence to a nonexplanatory usage are suggestions for an alternative approach to understanding problem solving proficiency. Essentially what is being suggested is not that intelligence be replaced by a new term but that it be relegated to a descriptive function applicable to certain kinds of behavior whose explanation can be attempted by the continued development of existing constructs which can better integrate data of the kind previously mentioned and direct attention to new avenues of research. As argued by McCandless (1952) and Rotter (1958) among others, the constructs of modern learning

theories would appear to be a likely point of departure for this endeavor.

Rather than regarding intelligence tests as measures of a single dimension (general intelligence) it is more in keeping with the evidence to interpret them as assessing culturally expected standards of academic preparedness involving a number of complexly interrelated variables (e.g., implicit reaction patterns, motives, habits, expectancies, reinforcements, etc.) to account for the results. Attempting to incorporate the varieties of socially learned behaviors operating in the testing situation within the domain of developing theories of learning has the salutary effect of replacing a notion which is essentially untestable with constructs which have been selected with a view to functioning in potentially fruitful theories.

Summary

An attempt has been made to question critically the validity of the term intelligence as an explanatory construct. It has been argued that the term intelligence carries with it essentially untestable and empirically unjustifiable connotations which preclude coherent and systematic organization of the awesome variety of activities ascribed to its domain.

An analysis of the antecedent and subsequent conditions defining intelligence reveals the following difficulties inherent in this concept:

1. The genetic model representing the antecedent conditions of intelligence is essentially untestable and not in accord with recent conceptual advances in both genetics and psychology.

2. The usefulness of postulating genetic antecedents for complex problem solving behaviors prior to a much more thorough understanding of how heredity and environment interact in a given case (a long and tortuous process at best) appears quite dubious both on theoretical and practical grounds. But more to the point are the more immediate possibilities of understanding and predicting intelligent behavior by continuing to pursue lines of investigation designed to reveal the situational conditions which facilitate or hinder the acquisition of various problem solving skills and the conditions which facilitate or hinder the expression of these skills in certain kinds of situations.

3. On the subsequent response side of the problem, the large majority of the evidence (primarily from factor analytic studies) indicates that a single dimension is not capable of subsuming the great varieties of behaviors now assumed to be related to intelligence.

4. The very generality of the term prohibits the isolation of a functionally related class of behaviors. Any attempt to establish a stable functional unity is further complicated by the definitional necessity of involving cultural value judgments in selecting a population of intelligent behaviors, judgments which vary from one culture to another and tend to change with time in a given culture.

5. Perhaps most crucial is the failure of intelligence theory to specify systematically the effect of situational variables on the behaviors it is purported to predict.

6. Finally, the arguments against the explanatory usage of intelligence suggest a fruitful approach to be the incorporation of the behavioral realm typically ascribed to intelligence within the confines of modern learning theory, particularly a social learning theory.

REFERENCES

Anastasi, Anne. Some implications of cultural factors for test construction. *Proc., 1949 Conf. Test. Probl., Educ. Test Serv.*, 1950, 13–17.

Anastasi, Anne. *Differential psychology*. New York: Macmillan, 1958. (a)

Anastasi, Anne. Heredity, environment, and the question "how?" *Psychol. Rev.*, 1958, **65**, 197–208. (b)

Anderson, J. E. The prediction of terminal intelligence from infant and preschool tests. *39th Yearbook, Nat Soc. Educ.*, 1940, Part I, 385–403.

Axline, Virginia. Mental deficiency—symptom or disease. *J. consult. Psychol.*, 1949, **13**, 313–327.

Bayley, Nancy. On the growth of intelligence. *Amer. Psychologist*, 1955, **10**, 805–818.

Beach, F. A., & Jaynes, J. Effects of early experience upon the behavior of animals. *Psychol. Bull.*, 1954, **51**, 239–263.

Beach, F. A. Characteristics of masculine sex drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1956, 1–32.

Carnap, R. The methodological character of theoretical concepts. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science*. Vol. I. Minneapolis: Univer. Minnesota Press, 1956, 38–76.

CLARKE, R. S., HERON, W., FETHERSTONHAUGH, M. L., FORGAYS, D. G., & HEBB, D. O. Individual differences in dogs: Preliminary report on the effects of early experience. *Canad. J. Psychol.*, 1951, **5**, 150–156.

COOMBS, A. W. Intelligence from a perceptual point of view. *J. abnorm. soc. Psychol.*, 1952, **47**, 662–673.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

EELLS, K., DAVIS, A., HAVIGHURST, R. J., HERRICK, V. E., & TYLER, R. *Intelligence and cultural differences*. Chicago: Univer. Chicago Press, 1951.

FROMM, E., & HARTMAN, L. *Intelligence: A dynamic approach*. New York: Doubleday, 1955.

FULLER, J. L. *Nature and nurture: A modern synthesis*. New York: Doubleday, 1954.

GORDON, E. M., & SARASON, S. B. The relationship between "test anxiety" and "other anxieties." *J. Pers.*, 1955, **23**, 317–323.

GUILFORD, J. P. The structure of intellect. *Psychol. Bull.*, 1956, **53**, 267–293.

HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

HUTT, M. L., & GIBBY, R. G. *The mentally retarded child*. Boston: Allyn & Bacon, 1958.

HUXLEY, J. *Heredity east and west*. New York: Henry Shuman, 1949.

JENNINGS, H. S. *The biological basis of human nature*. New York: Norton, 1930.

JESSOR, R., & HAMMOND, K. R. Construct validity and the Taylor Anxiety Scale. *Psychol. Bull.*, 1947, **54**, 161–170.

JESSOR, R. The problem of reductionism in psychology. *Psychol. Rev.*, 1958, **65**, 170–178.

LIVERANT, S. The use of Rotter's Social Learning Theory in developing a personality inventory. *Psychol. Monogr.*, 1958, **72**(2, Whole No. 445).

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

MANDLER, G., & SARASON, S. B. The effect of prior experience and subjective failure on the evocation of test anxiety. *J. Pers.*, 1953, **21**, 336–341.

MCCANDLESS, B. Environment and intelligence. *Amer. J. ment. Defic.*, 1952, **56**, 674–691.

MCCLELLAND, D. C. *Personality*. New York: Sloane, 1951.

MCGRAW, M. B. *Growth: A study of Johnny and Jimmy*. New York: Appleton-Century, 1935.

PEAK, HELEN. Problems of objective observation. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden, 1953, 243–299.

RAPAPORT, A. *Operational philosophy*. New York: Harper, 1953.

ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

ROTTER, J. B. A historical and theoretical analysis of some broad trends in clinical psychology. In S. Koch (Ed.), *Psychology: A study of science. Study II. Empirical substructure and relations with other sciences*. New York: McGraw-Hill, in press.

SARASON, S. B. *Psychological problems in mental deficiency*. New York: Harper, 1959.

SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810–817.

SHOFIELD, W. Critique of scatter and profile analysis of psychometric data. *J. clin. Psychol.*, 1952, **8**, 17–22.

SKINNER, B. F. Reinforcement today. *Amer. Psychologist*, 1958, **13**, 94–99.

SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Mental growth and personality development: A longitudinal study. *Monogr. Soc. Res. Child Developm.*, 1958, **23**(No. 68).

SPIKER, C. C., & MCCANDLESS, B. R. The concept of intelligence and the philosophy of science. *Psychol. Rev.*, 1954, **61**, 255–266.

STERN, C. *Principles of human genetics*. San Francisco: Freeman, 1952.

STONE, J. A critique of studies of infant isolation. *Child Developm.*, 1954, **25**, 9–20.

WECHSLER, D. *The measurement and appraisal of adult intelligence*. Baltimore: Williams & Wilkins, 1958.

(Received May 18, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

PERSONALITY DIFFERENCES AND LEVEL OF ASPIRATION

JUNE ELIZABETH CHANCE
University of North Carolina

A number of previous investigations have sought to study the relationship between personality differences and different patterns of level of aspiration. While many of these studies have found no stable relationships, a few (Cohen, 1954; Gruen, 1945; Klugman, 1948; Sears: 1940, 1941) find a recurrent relationship between "maladjustment" and particular patterns of goal setting. Specifically, two patterns appear to be more often characteristic of maladjusted subjects (*Ss*) than others, either very high (compensatory or wishful) or very low (protective) levels of aspiration. Cohen (1945), in discussing his results, asks what individual differences might explain why some *Ss* with strong feelings of self-rejection deal with the level of aspiration situation by making low, self-protective statements while others deal with the same situation by making high, unreal statements?

The present study concerns goal setting behavior as it relates to two independent dimensions of maladjustment in personality functioning. These two dimensions, designated A and R, were measured by two special MMPI scales devised by Welsh (1956). These scales, obtained by factor analytic methods, are uncorrelated with each other and may be used singly or in combination. While the scales are at present designated only as A and R, Welsh (1956) notes that high A scores are "related to disability of a dysthymic and dysphoric nature in which anxiety is prominent" while high R scores are "characterized by repression and denial." Preliminary work with psychiatric populations indicates that when the two scales are used together to classify *Ss* the resulting four groups are distinguishably different in psychiatric diagnosis (Welsh, 1956). Investigations with college students (Chance, 1957; Pepper, 1957; Saute, 1958) have likewise demonstrated differences in behavior between groups selected on the basis of position on both scales.

The present study was designed to investigate two hypotheses: (*a*) *Ss* scoring higher on A than on R (sensitizers)[1] will more characteristically employ a protective or low pattern of goal setting in a failure situation; (*b*) *Ss* higher on R than on A (repressors) will more often employ a compensatory or high pattern of goal setting when confronted with failure. These two hypotheses are consistent with a picture of the sensitizer as an individual who anticipates and fears failure and criticism and who deals with the situation by "admitting it first" and with a picture of the repressor as one who avoids discomfort by denying negative things about himself.

METHOD

Seven sections of elementary psychology courses were administered an adapted group form of the MMPI, containing the A and R scales, and a group level of aspiration measure during one of their regular class meetings. The group level of aspiration measure consisted of three sets of 15 anagrams and one set of 15 number series.

These four tasks were presented in booklet form with the following instructions:

> On the following pages you will be asked to take a number of short experimental tests. These are new tests which we are just trying out. Preliminary work indicates that they may be good tests of intelligence or general academic aptitude. We are also interested in how well people can predict their own test performance before taking a test; therefore, before each test you will asked to make a prediction of the score you think you are most likely to obtain.
>
> On each of the next three pages you will find 15 series of scrambled letters. If the letters are prop-

[1] The terms "sensitizers" and "repressors" will be used here for convenience. They refer only to the operational definitions given.

erly rearranged they will make a common English word. No proper names or trick words are included. Do any that you can; you need not work on them in order. You will have 2½ minutes to work on each set of anagrams.

When I say "Go" you may turn the page and begin. But before beginning, please write in the blank below the number of words you think you will be able to complete. On each of the series of anagrams the average college student is able to complete 8 of the 15 words. Nearly all students complete 4; a few students complete 12. Please do not turn the page until you are told to do so.

The instructions were read aloud by *E* while *Ss* followed them in their test booklets. *Ss* were asked to fill in identifying data on the front page as well as to make an estimate of their probable performance on the first set of anagrams. On a signal from *E*, all *Ss* turned to the second page and began working. They were told to stop at the end of 110 seconds. At the bottom of the second page they then recorded their estimates of their probable performances on the second set of anagrams. On a signal from *E* they turned to the third page and began work.

The procedure was repeated for the third and fourth pages; however, on the fourth page they were told:

On the next page you will find 15 number series. Four numbers are given and they are arranged according to some system. Figure out the system and then circle the one of the alternative numbers in the second column that fits the system. Three examples are given below.

The average college students gets 7 correct solutions out of 15 in 4½ minutes; nearly all complete 4, a few get 13. Please estimate in the blank below how many you think you will be able to complete. Do not turn the page until you are told to do so.

The time limit of 110 seconds for the sets of anagrams was so selected that the majority of *Ss*, if they set their aspiration close to the supposed group average of 8 correct solutions, could not succeed. (Although *Ss* were told that they were to be permitted a longer time to work than was actually allowed, no *S* questioned the imposed time limit.) For the number series they were actually allowed 4½ minutes to work. This longer time limit combined with the lower supposed "group average" was designed to let most *Ss* reach or exceed their aspiration estimate and hence to obtain some success at the end of the experiment.

Personality test and aspiration data were obtained from 147 *Ss*. In order to test the hypotheses regarding differences in aspiration behavior of sensitizers and repressors, *Ss* were selected from this larger group. Only male *Ss* were included in the experimental groups, since previous research has shown sex differences in level of aspiration behavior (Sheehan & Zelen, 1955; Sumner & Johnson, 1949).

A and R scores were translated into *T* scores by means of a table provided by Welsh (1956). Sensitizers were defined as those with A scores above 50, R scores below 50, and with at least a 10-point difference between the two scores. Repressors were similarly defined. For comparison purposes a control (A = R) group was also selected; controls obtained

TABLE 1

MEANS AND STANDARD DEVIATIONS OF PERFORMANCE SCORES, ASPIRATION ESTIMATES, AND PERFORMANCE-ASPIRATION DIFFERENCES

	Tasks							
	Anagrams 1		Anagrams 2		Anagrams 3		Number Series	
Groups	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Sensitizers (N = 17)								
Performance	5.88	2.15	4.53	1.58	3.82	2.21	7.35	2.06
Aspiration	8.00	1.61	6.41	1.82	5.24	1.22	6.88	1.37
Difference	−2.12	2.57	−1.88	3.14	−1.42	2.44	—	—
Repressors (N = 25)								
Performance	5.44	1.65	4.08	1.92	4.52	1.72	7.08	2.50
Aspiration	8.48	1.58	7.32	2.05	6.52	2.10	7.12	1.68
Difference	−3.04	2.31	−3.24	2.72	−2.00	3.08	—	—
Controls (N = 26)								
Performance	6.15	2.19	4.65	1.99	4.69	2.72	7.42	2.70
Aspiration	8.46	1.43	7.15	1.59	6.73	1.52	7.27	2.41
Difference	−2.31	2.83	−2.50	2.07	−2.04	2.33	—	—

A and R scores within the range from 45 to 55 and had no more than 5 points difference between their two scores. In addition no *S* was included in any group who had actually succeeded in meeting or exceeding his aspiration estimate on any of the three anagram tasks.

Results

Means and standard deviations of actual performances, aspiration estimates, and discrepancies between performance and aspiration on each task are given in Table 1. Mean differences between groups were evaluated by *t* tests. For each task (Anagrams 1, 2, and 3 and Number Series) the mean score of the sensitizers was compared with the mean score of the repressors; means of each of these groups were also compared with the mean of the control group. No differences between the groups in actual performance were significant. Likewise, the groups did not differ in size of discrepancies between performance and aspiration for the same task (which might be a measure indicative of intensity of failure). However, the mean aspiration estimate of the sensitizers for the third set of anagrams was significantly different from either the mean of the repressors or the mean of the control group. Values of *t* were, respectively, 2.25 ($p < .02 > .01$, df 40) and 3.43 ($p < .01 > .001$, df 41). None of the other means of aspiration estimates differed significantly among the groups.

Table 2 presents means and standard deviations of *d* scores, i.e., the discrepancy between aspiration estimate and performance on the previous trial. Mean differences were tested in the same manner as for the data in Table 1. None of the mean differences between groups in the first *d* scores are significant, although the sensitizers had scores lower than the repressors as predicted ($t = 1.84$, $p < .10 > .05$, df 40). The mean difference between sensitizers and repressors in the second *d* score is highly significant ($t = 2.94$, $p < .01 > .001$, df 40). The same mean of the sensitizers approaches being significantly different from that of the control group ($t = 1.82$, $p < .10 > .05$, df 41). None of the comparisons of means of the third *d* score, which was based upon aspiration for the number series, were significant.

TABLE 2

Means and Standard Deviations of *d* Scores (Aspiration Estimate Minus Performance on Previous Trial)

	Scores					
	$d_1 = A_2 - P_1$		$d_2 = A_3 - P_2$		$d_3 = A_4 - P_3$	
Groups	*M*	*SD*	*M*	*SD*	*M*	*SD*
Sensitizers	.53	2.53	.65	1.68	3.06	2.97
Repressors	1.96	2.39	2.44	2.06	2.60	1.77
Controls	1.00	2.47	2.08	1.64	2.58	2.98

Examination of Tables 1 and 2 reveals that all groups begin the experiment with roughly equivalent levels of aspiration, perform in about the same manner, and fail to about the same extent. All groups tend to shift their aspiration estimates downward over the three trials of anagrams and to raise them again when the number series is introduced. However, the sensitizers as a group tend to shift their aspiration levels downward to a more marked degree than the other two groups, thus keeping the discrepancy between their estimate for any trial and their previous performance relatively small as compared with either the repressors or the control group. These data might be considered to confirm tentatively the hypothesis that sensitizers would be characterized by the use of a protective pattern of goal setting in a failure situation. The hypothesis that repressors would be characterized by a high, compensatory pattern fares less well. Although the repressors as a group do tend to keep the discrepancy between their last performance and next aspiration relatively larger than do the sensitizers, their *d* scores are not larger than those of an equal tendency control group.

Discussion

The result, while suggesting the sensitizer is prone to adopt a protective pattern of goal setting in a failure situation, does not find any clear indication that the repressor adopts a compensatory pattern. Inspection of the data also failed to confirm an alternative hypothesis that repressors might be utilizing two or more patterns rather than just one. Both repressors and controls in making esti-

mates either retained the estimate given for the previous trial or else lowered their old estimates slightly, i.e., there were no marked differences between repressors and controls in their responses to this experimental situation.

Although the original intention of the study was to use the A and R scales as two dimensions of maladjustment, the choice of operational definitions for sensitizers and repressors would argue that few, if any, *Ss* included in either group could legitimately be characterized as markedly deviant. Clearer differences between sensitizers and repressors and between both these groups and the control group might have been obtained if more extremely deviant *Ss* had been available.

In a carefully executed study, Miller (1951) investigated level of aspiration in psychiatric patients. Among his several criteria for inclusion of patients in one of the four diagnostic groups studied were patterns of scores from the clinical scales of the MMPI. Welsh's (1956) demonstrated relationships between the A-R classification and profile configurations on the MMPI suggests that the findings of the present study utilizing a "nondeviant" population may be profitably compared with Miller's findings from a "deviant" population. Miller found that neurasthenics produced on the average the highest positive *d* score, that both paranoid schizophrenics and a group of nonpsychiatric hospital patients produced *d* scores just slightly above zero, and that character disorders and conversion hysterics produced slightly negative *d* scores. Welsh found that elevation on *D*, *Pt*, and *Sc* were prominent in the profiles of high A/low R *Ss* and that psychiatric diagnoses of severe psychoneurosis, anxiety state, and paranoid schizophrenia occurred most often in this group. High R/low A *Ss* were more likely to show elevations on *Hs* and *Hy* and a marked lack of elevations on the psychotic scales; they were more likely in a psychiatric population to obtain diagnoses of mild psychoneurosis, hysteria, and conversion reaction. Examination of Miller's definition of neurasthenia (elevation on *Hs*, *D*, and to a lesser extent on *Hy;* lack of elevation on *P*, *Sc*, and *Ma*) indicates a fair degree of correspondence to Welsh's diagnostic expectations for the high R/low A group, and, as in the present study, this is the group found to have the highest positive *d* scores. Similarly, Miller's diagnoses of character disorder and paranoid schizophrenia correspond rather clearly to the high A/low R scoring group described by Welsh. And again the observations of both studies are congruent; both groups show a low average *d* score, implying a self-protective orientation toward goal setting.

There are, however, some discrepancies in the findings of the two studies. Miller's normals set goals more like his paranoid schizophrenics and character disorders, while in the present results normals set goals more like repressors than sensitizers. The second departure in the findings is the negative *d* score of Miller's conversion hysterics. Following Welsh, we would expect these *Ss* to behave like the repressor group, obtaining positive rather than negative *d* scores. Miller's criteria for the diagnosis of conversion hysteria seem, however, to have been somewhat more stringent than those employed in the construction and validation of the *Hy* scale, which may account in part for the discrepancy of results. The over-all degree of correspondence between the present findings and Miller's lends considerable support to the general idea that characteristic patterns of goal settings and defending against failure are a significant aspect of different personality organizations all along the continuum of adjustment.

The present study also differs in several respects from many previous studies. One important difference is that *Ss* were subjected to consistent failure throughout the first three trials of the experiment; consequently, the obtained results could not be generalized to anything other than a failure situation. The influence of personality trends on goal setting might be expected to operate differently in situations where *Ss* succeed or where success and failure are interspersed. This experiment also utilized a relatively short series of trials; the character of the results might change if *Ss* were subjected to more prolonged failure.

Summary

Groups of college students classified as "sensitizers" or "repressors" by means of the Welsh A and R scales (MMPI) were com-

pared in their performance on a group level of aspiration measure. The situation was arranged so that *Ss* experienced failure on the LOA tasks. Sensitizers employed a goal setting pattern closer to their real previous performance than did repressors or an equal tendency control group (A = R). Repressors did not show an hypothesized tendency to set high or compensatory goals. Limitations of the findings are discussed.

REFERENCES

CHANCE, JUNE E. Some correlates of affective tone of early memories. *J. consult. Psychol.*, 1957, **21**, 203–205.

COHEN, L. D. Level of aspiration behavior and feelings of adequacy and self acceptance. *J. abnorm. soc. Psychol.*, 1954, **49**, 84–86.

GRUEN, EMILY W. Level of aspiration in relation to personality factors in adolescents. *Child Develpm.*, 1945, **16**, 181–188.

KLUGMAN, S. F. Emotional stability and level of aspiration. *J. gen. Psychol.*, 1948, **38**, 101–119.

PEPPER, L. J. The relationship between Welsh Figure Preference Test responses and indices of anxiety and repression. Unpublished master's thesis, Univer. of North Carolina, 1957.

MILLER, D. R. Responses of psychiatric patients to threat of failure. *J. abnorm. soc. Psychol.*, 1951, **46**, 378–387.

SAUTE, LOIS C. Judge personality and favorability biases in interpersonal predictions. Unpublished doctoral dissertation, Univer. of North Carolina, 1958.

SEARS, PAULINE S. Levels of aspiration in academically successful and unsuccessful children. *J. abnorm. soc. Psychol.*, 1940, **35**, 498–536.

SEARS, PAULINE S. Level of aspiration in relation to some variables of personality: Clinical studies. *J. soc. Psychol.*, 1941, **14**, 311–336.

SHEEHAN, J. G., & ZELEN, S. L. Level of aspiration in stutterers and nonstutterers. *J. abnorm. soc. Psychol.*, 1955, **51**, 83–86.

SUMNER, F. C., & JOHNSON, E. E. Sex differences in levels of aspiration and self estimates of performance in a classroom situation. *J. Psychol.*, 1949, **27**, 483–490.

WELSH, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.

(Received February 13, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

MANIFESTATIONS OF HOSTILITY IN NEURODERMATITIS [1]

M. MICHAEL KLABER [2]
Teachers College, Columbia University

Neurodermatitis is a relatively common chronic skin disorder, characterized by intensely pruritic excoriations, and is usually attributed to psychogenic factors. The nature of these psychological disturbances is described as centering around the suppression of intense hostile impulses (Dunbar, 1954; Obermayer, 1955). Neurodermatitis patients thus present an opportunity to investigate a specific personality pattern in a relatively objectively classified population. The purpose of this research was to test by an experiment the conclusions obtained through clinical observations, and to contribute to a better understanding of the disorder.

Early studies using rating scales and personality inventories (Allerhand, Gough, & Grais, 1950; Greenhill & Finesinger, 1942; Lynch, Hinckley, & Cowan, 1945) unanimously reported a higher incidence of hostility in neurodermatitis patients than in various control groups. None of these investigators, however, reported a cross-validation of their findings.

Several projective studies are reported in the literature (Cleveland & Fisher, 1956; Fiske, 1953; Fiske & Obermayer, 1954; Geist, 1957; Kepecs, Rabin, & Robin, 1951; Levy, 1952). All investigators concurred with the finding that neurodermatitis patients manifest more hostility than do normal persons on projective tasks. Little is known, however, of the relationship between covert hostility as manifested in projective protocols and overt expression of aggression. In an effort to correlate various levels of assessment of behavior, Leary (1957, pp. 395–396) constructed a multilevel personality battery. Reporting on skin patients he described his findings: "The multilevel pattern is that of a sweet, responsible facade with underlying rage and bitterness . . . they exhibit an intense conflict between this bland exterior and sadistic rage." Unfortunately, no control group was used for comparison with the skin patients.

An outstanding therapeutic experiment conducted by Seitz (1953) emphasized the suppression of hostility in dermatitis patients and based a therapeutic procedure on this characteristic. Reporting on a series of 25 patients, he arrived at the conclusion that the alleviation of skin eruptions could be achieved through a short series of therapeutic interviews focusing only on hostility and aggressive impulses. In a period of 10 sessions he explored possible sources of annoyance to the patient and encouraged him to talk about them. Only 13 of the 25 patients remained in therapy; 12 of the 13 were completely relieved of symptoms.

Summarizing the present state of knowledge, the neurodermatitis patient may be described as a generally hostile person who expends great effort on successful control of his impulses. Outwardly he appears calm and collected, yet covertly he harbors deep resentment. In psychotherapy such a patient will readily respond to favorable interpersonal conditions by verbalizing a great deal of hostility. This research was designed to test these formulations.

[1] This study is based on a doctoral dissertation submitted in partial fulfillment of the requirements for the PhD degree at Teachers College, Columbia University. Thanks are extended to Joel R. Davitz, Paul E. Eiserer, and especially to Laurance F. Shaffer, who was of invaluable assistance during all stages of this study. The author is indebted to A. C. Carr and G. S. Baroff for serving as judges.

[2] Now at the Institute for the Crippled and Disabled.

Method

The instruments were chosen to reflect the central issues of the research: They were (*a*) a Manifest Scale of Hostility to test the hypothesis that neurodermatitis patients do not see themselves as more aggressive than normal persons; (*b*) seven Thematic Apperception Test (TAT) cards selected to elicit hostility and thus to tap the covert level of this specific behavior; and (*c*) a Sentence Formation Task (SFT) on which sentences were constructed and which could be reinforced. It was postulated that the SFT best reflects the therapeutic situation and thus represented a laboratory analogue to Seitz's method.

The Manifest Scale of Hostility

A manifest scale on which the subject (*S*) evaluated himself with respect to hostile behavior was chosen for comparing overt hostility. Inasmuch as the *Ne* scale (Allerhand et al., 1950) had been especially designed for use with neurodermatitis patients and contained a number of items dealing with the central issue of this research, its use seemed appropriate.

Twenty-four items from the *Ne* scale were included in the instrument, supplemented by four additional items in order to obtain greater reliability. The score was the sum of the items judged by two psychologists to pertain to hostility. In such a manner 15 items were included in the final score.[3] In spite of the added items in an effort to obtain a reliable measure, the index of split-half reliability was low ($r = .39$).

The Thematic Apperception Test

For the purpose of obtaining a covert hostility score, a series of seven TAT cards was administered in the standard manner. Cards 3 BM, 6 BM, 8 BM, 12 M, 13 MF, 18 GF, and 18 BM were used. The stories obtained were independently analyzed by two clinical psychologists according to the scoring instructions developed by Fisher and Hinds (1951). It was emphasized that scoring should not involve interpretation, but mere classification of content. Seven categories of hostility were classified from story content: (*a*) openly hostile stories, (*b*) verbally expressed hostility, (*c*) rationalized hostility and suicide, (*d*) illness and accidents, (*e*) ambivalently expresed hostility, (*f*) neutral content, and (*g*) positive or favorable content. Interjudge reliability for the scoring categories ranged from .79 to .85 on the different scales. In those instances where judges disagreed, a third clinician was consulted to resolve the issue.

The scores were weighted in the following manner: a score of 3 was assigned to Category *a* (openly hostile stories), a score of 2 to Categories *b* and *c* (verbal expressions of hostility and rationalized hostility), a score of 1 to Categories *d* and *e* (illness, accidents, and ambivalently expressed hostility), a score of −1 to neutral content (Category *f*), and of −2 to positive content (Category *g*). The final score was the algebraic sum of the scores of the seven stories for one individual.

The Sentence Formation Task

While some previous studies used overt and covert measures, few, if any, attempted to measure the process of psychotherapy itself. Psychotherapy has been viewed as a process of selective reinforcement by the therapist (e.g., Shaffer, 1947). Inasmuch as the elicitation of hostile responses is central to the psychologic treatment of neurodermatitis patients, it was believed that verbal reinforcement of hostile responses is analogous to the therapeutic situation. A recent review by Krasner (1958) summarized the available research in this area.

In order to elicit reinforcable responses, a sentence completion task was constructed. On each of 74 $3'' \times 5''$ cards four verbs were typed, one of which was of hostile stimulus value. The *S* was presented with one card at a time and instructed to construct one sentence using any one of the four verbs presented to him.

The hostility value of a verb was determined by two clinical psychologists. Each was given a list of 82 verbs selected for possible hostility value, and was asked to indicate which of the verbs would be termed "hostile" when used in a sentence. Both judges agreed on 74 verbs, and these were used in the construction of the instrument.

Three scores were obtained for each individual consisting of the number of verbs of hostile stimulus used in each of the three conditioning periods: operant, reinforced, and extinction. A split-half reliability coefficient was computed for each of the three experimental conditions. The correlations for hostile responses were .60, .62, and .77, respectively.

Procedure

The manifest scale was administered in a mimeographed form on which the patient had to indicate, by circling either T or F, whether a statement applied to him or not.

The seven TAT cards were administered in the standard manner.

The Sentence Formation Task was introduced with the following statement: "I want to see how good you are at making sentences. I am going to show you some cards, each of which has four verbs. You can choose to make a sentence with any of the verbs." The first 30 cards were presented without any comment in order to establish an operant level of response. On the next 30 cards, whenever the *S* chose the hostile verb to make a sentence, he was reinforced by the experimenter who said, "yes," "good," "fine," "uh-huh," or "hmm-mm," as the oc-

[3] After the inception of the research Buss and Durkee published a Hostility-Guilt inventory which may have been particularly useful in this study. The publication came too late, however, to be incorporated in the research design (Buss & Durkee, 1957).

casion warranted. The presentation of the last 14 cards was regarded as an extinction period and no reinforcement was administered. The cards were presented in a random order.

The testing session required from 45 to 60 minutes.

Subjects

The *Ss* of the investigation were 40 patients referred through the Department of Dermatology of the Presbyterian Hospital and the College of Physicians and Surgeons of Columbia University as well as from one private dermatologist.[4] Three experimental *Ss* were obtained through the latter source, while the rest of the neurodermatitis patients, as well as all the controls, through the former.

In order to control for the psychological effects of skin disorder per se, all control *Ss* were derived from the same general population as the experimental group, except for their diagnosis. Only patients afflicted with clearly nonpsychogenic disorders (e.g., contact dermatitis, lupus, etc.) were accepted. The experimental *Ss* were diagnosed independently by a board-certified dermatologist or by the staff of the clinic.

All *Ss* were white, over 18 years of age, and were born or had received their schooling in this country. The experimental group consisted of 5 males and 15 females, and the control *Ss* were 8 males and 12 females. No significant differences were found between the experimental and control groups with respect to age (the means were 38.75 and 38.10, respectively) and years of school attended (means 12.35 and 12.85, respectively).

Results and Conclusions

Overt Hostility

No significant differences between the neurodermatitis patients and the control group were found ($t = 1.02$ with 38 df). One must bear in mind, however, that the reliability coefficient of the scale used was only .39, so that the acceptance of the null hypothesis applied only to the measure used. Further research with a more reliable scale would throw additional light on the matter at hand.

Covert Hostility

It had been hypothesized that neurodermatitis patients show greater hostility than the control group on instruments tapping covert personality factors, such as measured by the TAT. This hypothesis is borne out by the results of the research. Table 1 shows both significantly higher mean hostility scores and greater variance for the experimental *Ss* in comparison with the dermatology control group. These results show that neurodermatitis patients emit both more hostile responses and are as a group more variable on a projective test designed to tap covert personality variables.

TABLE 1

Comparisons of Means and Standard Deviations on TAT Scores

Statistic	ND Patients	Control	Statistics of Comparison	p
Mean	6.7	3.3	$t = 2.83$	.01
SD	5.1	2.8	$F = 3.22$	.02

Response to Reinforcement

From the reports of psychotherapists, it became evident that neurodermatitis patients respond favorably to verbal reinforcement of hostile responses. It was therefore hypothesized that such patients manifest different response frequencies of verbal hostile behavior under experimental reinforcement conditions than normal persons.

Three measures were computed for purposes of inter- and intragroup comparisons,

TABLE 2

Comparisons of Means and Standard Deviations of Hostility Scores on SFT

(In percent possible score)

Experimental Condition	ND Patients	Controls	Statistics of Comparison	p
Operant				
Mean	10.8	14.3	$t = 2.04$	.05
SD	5.0	5.6	$F = 1.26$	NS
Reinforced				
Mean	18.2	18.5	$t = 0.83$	NS
SD	9.4	8.6	$F = 1.19$	NS
Extinction				
Mean	18.5	18.5	$t = 0.00$	NS
SD	9.8	9.2	$F = 1.13$	NS

[4] The author is indebted to Carl T. Nelson, Executive Officer of the Department of Dermatology, for his permission to use the *Ss* and for his active cooperation in the study, and to D. P. Torre and J. McCarthy for their assistance and cooperation.

corresponding to operant, conditioning, and extinction periods. As can be seen from Table 2, neurodermatitis patients show a significantly lower hostile response frequency on the SFT under nonreinforced conditions than control *Ss*. Under favorable (i.e., reinforced) conditions, however, they emit no fewer hostile responses than the control group.

As it was originally stated, the hypothesis with respect to increased verbal response rates of neurodermatitis patients under favorable conditions was not substantiated. These patients differ from the control group not in response to reinforcement, but in their base level of verbal hostility. Reinforcement enables these persons to achieve a normal rate of emission of verbal hostile responses. The results are consistent with the clinical observations.

Hostility and Neurodermatitis

Certain conclusions with respect to the manifestations of hostility in neurodermatitis patients may be drawn from the test battery used in this research. Assuming that the Manifest Scale of Hostility measures overt, and the TAT covert, expressions of aggressive impulses, the neurodermatitis patients emerge as persons with a covert hostility drive which they do not manifest overtly. This conclusion is substantiated by their low operant level of emittance of hostile verbs. The spontaneous, unreinforced response may constitute a sample of the verbal hostility rate of the person in everyday life. A low operant level thus reflects infrequent use of hostile words and consequently placid verbal behavior.

Discussion

The results of this study demonstrate the feasibility of constructing a test battery specifically designed to measure overt and covert levels of a specific personality dimension, as well as the verbal emission of certain response categories. Such batteries may, in the future, be used to investigate the nature of the psychotherapeutic process and may contribute to a better understanding of the verbal interaction between therapist and client.

The research study concerned itself with a specific medical-diagnostic group of patients and confirmed impressions formulated by clinical workers with respect to one character dimension (hostility), but was not designed to test all hypotheses concerning the etiology of the disorder. It is entirely possible, for example, that other personality characteristics, such as anxiety, are concomitants of the neurodermatitis syndrome. Furthermore, it is possible and quite likely that disturbances in the psychologic sphere are necessary contributing factors to the etiology of the disorder, but that an organic predisposition constitutes a prerequisite for the formation of the syndrome.

The findings of the study are limited by the deficiencies of the instruments used. The Manifest Scale of Hostility in particular was found to be so low in reliability that the acceptance of the null hypothesis with respect to the two groups may well have resulted from the inadequacy of the scale rather than from a real absence of difference in response pattern. Inasmuch as available research presents conflicting evidence with respect to overt and covert relationships, the data of this study are insufficient to demonstrate conclusively that the hostility manifested in neurodermatitis patients is demonstrable on a purely covert level. Taken at face value, the results obtained on the TAT and the SFT are consistent with theories postulating an inverse relationship with respect to overt and covert hostility patterns. The lower unreinforced use of hostile verbs on the latter measure can be regarded as a behavioral measure of verbal hostility.

The experimental procedure in the research was designed in such a manner as to permit future applications to research with the process of psychotherapy. The SFT appears suited to simulate an experimental psychotherapeutic relationship.

Summary

Substantial clinical evidence suggests that patients afflicted with neurodermatitis suffer from a personality disturbance in which the suppression of intense hostility is a central feature. This research was designed to obtain empirical evidence concerning the manifestations of hostility on overt and covert levels in such patients.

It was hypothesized that neurodermatitis patients (*a*) evidence no more hostility than control *Ss* suffering from nonpsychosomatic skin diseases on an instrument tapping overt behavior, (*b*) that neurodermatitis patients manifest more hostility on a covert measure of hostility than the control *Ss*, and, finally, (*c*) that neurodermatitis patients differ in verbal hostile responses on a conditioning task.

The results were consistent with all three hypotheses. No difference between the two groups was found on the Manifest Scale of Hostility, but the low reliability of the scale does not permit more than a tentative acceptance of the null hypothesis. Differences on the TAT, as rated by two independent clinicians, were significant at the .01 level with respect to both higher mean and larger variance in favor of the neurodermatitis group. On the Sentence Formation Task the control patients emitted significantly more hostile responses on the operant base level than the experimental *Ss*. The experimental *Ss* increased their hostile responses under reinforcement conditions, reaching a level equal to that of the control *Ss*. Some of the implications of the methodology with respect to future research were discussed.

REFERENCES

ALLERHAND, M. E., GOUGH, H. G., & GRAIS, M. C. Personality factors in neurodermatitis: A preliminary report. *Psychosom. Med.*, 1950, **12**, 387–390.

BUSS, A. H., & DURKEE, ANN. An inventory for assessing different kinds of hostility. *J. consult. Psychol.*, 1957, **21**, 343–349.

CLEVELAND, S. E., & FISHER, S. Psychological factors in the neurodermatoses. *Psychosom. Med.*, 1956, **18**, 209–220.

DUNBAR, FLANDERS. *Emotions and bodily changes.* (4th ed.) New York: Columbia Univer. Press, 1954.

FISHER, S., & HINDS, EDITH. The organization of hostility controls in various personality structures. *Genet. psychol. Monogr.*, 1951, **44**, 3–68.

FISKE, C. E. Personality and emotional factors in chronic disseminated neurodermatitis. Unpublished doctoral dissertation, Univer. of California, Los Angeles, 1953.

FISKE, C. E., & OBERMAYER, M. E. Personality and emotional factors in chronic neurodermatitis. *Arch. Derm. Syph., N. Y.*, 1954, **61**, 261–268.

GEIST, H. Emotional aspects of dermatitis. *J. clin. exp. Psychopath.*, 1957, **18**, 87–93.

GREENHILL, M. H., & FINESINGER, J. E. Neurotic symptoms and emotional factors in atopic dermatitis. *Arch. Derm. Syph., N. Y.*, 1942, **46**, 187–200.

KEPECS, J. G., RABIN, A., & ROBIN, M. Atopic dermatitis. *Psychosom. Med.*, 1951, **13**, 1–9.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

LEARY, T. *Interpersonal diagnosis of personality: A functional theory of personality evaluation.* New York: Ronald, 1957.

LEVY, RUTH. Rorschach pattern in neurodermatitis. *Psychosom. Med.*, 1952, **14**, 41–49.

LYNCH, F. W., HINCKLEY, R. G., & COWAN, D. W. Psychosomatic studies in dermatology: B. Psychobiologic studies of patients with atopic eczema (disseminated neurodermatitis). *Arch. Derm. Syph., N. Y.*, 1945, **51**, 251–257.

OBERMAYER, M. E. *Psychocutaneous medicine.* Springfield, Ill.: Charles C Thomas, 1955.

SHAFFER, L. F. The problem of psychotherapy. *Amer. Psychologist*, 1947, **2**, 459–467.

SEITZ, P. F. Dynamically-oriented brief psychotherapy: Psychocutaneous excoriation syndromes. *Psychosom. Med.*, 1953, **15**, 200–242.

(Received February 16, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

A COMPARISON OF CHANGES IN PSYCHONEUROTIC PATIENTS DURING MATCHED PERIODS OF THERAPY AND NO THERAPY [1]

ROSALIND DYMOND CARTWRIGHT AND JOHN L. VOGEL [2]

University of Chicago

Despite a very large increase in the number of published research reports on the effects of psychotherapy, the more basic question as to its effectiveness seems still to remain unanswered. This question has been framed most often as one of the effectiveness of therapy versus spontaneous remission. It has been dealt with in these terms in a review by Eysenck (1952) and in recent research studies by Barron and Leary (1955), and Dymond (1955). Barron and Leary concluded from their careful study of MMPI changes that therapy patients did not improve significantly more than did waiting list controls. Cartwright (1956), in a note on their paper, pointed out that though therapy patients did not *improve* more, there was significantly more *change* with regular, formally defined psychotherapy than without it. This was indicated by the difference in the variances of the changes for the treated and untreated groups. Dymond (1955) reported a comparison of improvements made by a small sample of waiting list "improvers" and therapy "improvers." The analysis strongly suggested the improvement of the two groups to be different in kind.

Although neither study gives any definitive answer to the basic question, each supplies some evidence of therapy having some measurable impact not found in its absence. The study to be reported here is another of this kind: it appears to help clarify the argument that therapy does have some special effectiveness with psychoneurotic patients by specifying the conditions of the effectiveness more fully.

[1] This investigation was supported by a research grant from the Ford Foundation to the University of Chicago Counseling Center.

[2] Now at University of Washington.

HYPOTHESES

The major hypothesis can be stated most generally as:

1. There will be more change in measured adjustment for a self-referred group of psychoneurotic patients during a period of time in which they are meeting regularly with a therapist than during a period of time equal to it during which no regular formally constituted therapy is taking place.

Several subhypotheses will also be investigated. If it is supposed that therapy has more effect than no therapy, then:

2. Longer therapy periods will be associated with more improvement than shorter therapy periods.

3. Longer waiting time periods will not be associated with more improvement than shorter waiting periods.

4. Patients seen by experienced therapists will improve more in a therapy period than those seen by inexperienced therapists.

METHOD

Design and Subjects

Thirty subjects (*Ss*) who applied to the University of Chicago Counseling Center for therapy were asked to participate in a research study which would involve testing at four points in time: (*a*) immediately upon their agreement to participate (Pre-wait test), (*b*) immediately before beginning counseling (Pre-therapy test), (*c*) sometime during the course of their therapy at a point exactly equal in time to the length of time between Test 1 and Test 2 (In-therapy test), and (*d*) when therapy had been completed (Post-therapy test).

Of the 30 *Ss* who were initially tested, 8 remained

TABLE 1

Q ADJUSTMENT AND TAT SCORES OVER MATCHED TIME INTERVALS IN 22 EXPERIMENTAL SUBJECTS

	Q Adjustment Scores			TAT Mental Health		
	Pre-Wait	Pre-Therapy	In-Therapy	Pre-Wait	Pre-Therapy	In-Therapy
Mean	36.41	35.73	39.23	10.09	11.50	11.95
SD	10.25	12.77	10.11	2.18	3.04	2.95

in therapy for fewer than six interviews. These were called attrition cases and treated separately. The experimental sample, then, consisted of 22 cases: 11 males and 11 females who stayed in therapy for a mean of 33.4 hours, the range being 6 to 97 hours. There were 11 students and 11 nonstudents in the group. The mean age was 27 years and the range 18 to 41 years. The sample, though random in selection, was quite representative of the population regularly seen at this clinic.

In this study the waiting period was not fixed but varied according to the reality of therapist availability. Actually the mean waiting time for the experimental sample was 8.8 weeks and the range 3 to 24 weeks.[3]

Two test instruments were used: a self-descriptive test—the Butler and Haigh (1954) Q sort—from which the Q adjustment score (Dymond, 1954a) was computed, and a projective—the TAT—from which a mental health rating (Dymond, 1954b) was derived.[4]

There were 19 client-centered therapists involved in the study, 10 of whom were classed as experienced and 9 as inexperienced. An experienced therapist was defined as one who had handled six or more therapy cases previous to the present research case. The mean number of previous cases for the experienced group was 25.8 and the range 6 to 60; for the inexperienced group the mean was 1 and the range 0 to 5.

[3] A table giving all the basic data for this study, the distributions of time waiting, experience of therapist, and test scores at each testing point was eliminated at the suggestion of the editor to conserve space, but is available on request from the senior author.

[4] The mental health ratings were all made by the senior author. All identifying material both as to subject and time of the test was first removed. These were then rated on the same seven-point scale reported in previous studies (Dymond, 1954a, 1955). For those interested in comparing these results to those reported earlier, it should be noted that the scale has been expanded to score the additional plus and minus signs often added to the basic number rating. Thus the old scale, which ran from 1 = severe disturbance bordering on psychotic or psychotic to 7 = well integrated, happy person, socially effective, is now a 21-point scale covering the same range.

RESULTS

Table 1 gives the basic data for the matched time comparisons. It should be noted that the third test was at an In-therapy point which for some cases was close to the end of their therapy, but for others very far from it. For this reason a directional hypothesis would be a particularly stringent one. A more conservative hypothesis would be that employed here of greater change.

The major hypothesis (that there will be more change in a therapy period than in a No-therapy period) was tested as outlined below.

Let the change scores between Pre-wait and Pre-therapy be called W, and the change scores between Pre-therapy and the In-therapy point be called T. The same procedure was followed for both W and T scores.

(*a*) The mean W score was first computed.

(*b*) W scores were transformed by subtracting the mean as a constant and disregarding the sign. These transforms were called W'. Thus these transforms represent the absolute deviation of an individual W score from the mean W score.

(*c*) The mean of the transforms was then obtained.

(*d*) The difference between mean W' and mean T' was then computed, using t for matched groups.

For the Q adjustment score the t between W' and T' was 1.925, significant at about the .05 level. The t between the W' and T' scores for the TAT mental health ratings was .643, a nonsignificant result. Thus, as measured by the Q score, adjustment changes, regardless of direction, were significantly greater during a therapy period than during a No-therapy period. This was not true for the adjustment changes as measured by the TAT. It would appear, then, that the basic hypothesis—that

TABLE 2

LENGTH OF TIME WAITING AND DIRECTION OF CHANGE

Waiting Time	Q Adjustment Change		TAT Change	
	Positive (+)	No Change or Negative (0−)	Positive (+)	No Change or Negative (0−)
Long (8–24 weeks mean = 12.4)	6	6	9	3
Short (3–7 weeks mean = 4.5)	5	5	3	7
		NS $\chi^2 = 4.46$*		

* Significant at the .05 level.

therapy has more impact to change patients than an equal time period without therapy—is upheld for adjustment as measured from a self-descriptive test but not for adjustment as measured from a projective. It is interesting to note that the results found by Cartwright (1956) from the Barron and Leary (1955) study were also based on a self-descriptive type of test (the MMPI).[5]

Testing Hypotheses 2 and 3

Hypotheses 2 and 3 were concerned with the effects of the length of waiting and therapy time and their relation to change on the two tests. A chi square analysis was used.

Table 2 shows the direction of change on the Q adjustment score to be unrelated to the length of time without therapy. However, Table 2 shows that greater *waiting* time is related to more positive change on the TAT. The chi square analysis for the relation of direction of change to long and short *therapy* periods showed no relation for either test. The Table 2 result for the TAT is directly opposite to that hypothesized and raises the question: Is the direction of change while waiting related to the original status on the scores? Do high scores tend to decrease and low scores tend to increase?

The analysis of Q adjustment score showed no relationship between original status on this test and direction of change while waiting nor any relationship of Pre-therapy status and change while in therapy. For the TAT, a trend appeared for those with higher TAT scores at Pre-wait to deteriorate and those with lower scores to improve while waiting. This trend became significant in Table 3 with Pre-therapy high scorers decreasing in health on the TAT scale and low scorers improving during a therapy period.

To clarify some of the interrelationships of

TABLE 3

RELATION OF PRE-THERAPY ADJUSTMENT TO THE DIRECTION OF CHANGE WHILE IN-THERAPY

(Change while in Therapy)

	Q Adjustment			TAT		
		+	0−		+	0−
Pre-Therapy status	High 36+	5	6	High 12+	2	9
	Low 35−	8	3	Low 11−	7	4
		NS $\chi^2 = 4.70$*				

* Significant at .05 level.

[5] On both tests the *S* describes himself, using the given items which are then scored in terms of his correspondence to a criterion. For the MMPI this is the actual performance of the various diagnostic groups used in the standardization. For the Q adjustment score it is correspondence to a clinicians' standard of adjustment (Dymond, 1954b).

TABLE 4

EFFECT OF PRE-WAIT STATUS ON DIRECTION OF CHANGE AND LENGTH OF WAIT (TAT)

(Direction of Change on TAT while Waiting)

	+	0−	Total
High 10+			
Pre-Wait status			
Long Waits 8–24 Weeks	3	3	6
Short Waits 3–7 Weeks	1	4	5
Low 9−			
Pre-Wait status			
Long Waits	6	0	6
Short Waits	2	3	5
	12	10	22

the three variables—length of time (waiting or In-therapy), original status, and direction of change—a two-by-two-by-two analysis was used.[6] The first of these was designed to answer the question "Does the original Pre-wait status on the TAT affect the association between the length of waiting time and direction of change on this measure?" Separate two-by-two tables were set up to look at the direction of change positive as against no change or negative change for long wait and short wait *Ss*, one for those whose original status on the TAT was high and the second for those whose Pre-wait status was low.

These two tables were then combined in terms of number of cases on the diagonal and number off the diagonal. The difference between the proportions on the diagonal for high and low original status was tested for significance using the Fisher and Yates exact method. In this case, the difference between the proportions 7/11 and 9/11 is not significant. The answer to the original question, then, is that long wait *Ss* improve and short wait *Ss* do not improve or deteriorate on the TAT irrespective of their original status.

A second analysis of this type was done to determine if deterioration of the high Pre-therapy TAT scorers and improvement of the low Pre-therapy scorers during a therapy period was affected by the length of the time in therapy. The proportion on the diagonal for the long therapy group in this case was 7/10 and not significantly different from the proportion on the diagonal for the short therapy group 9/12. Thus the relationship found in Table 3 holds: those with high TAT scores before therapy begins tend to deteriorate or show no change, while those with initially low scores tend to improve these scores regardless of the length of time involved in the In-therapy period. It can now be stated in relation to Hypothesis 2 that longer therapy is not related to more improvement on either test. Hypothesis 3 that longer waiting would not be related to more improvement is confirmed for the Q adjustment score, but not for the TAT, where longer waiting was found related to more improvement regardless of original status.

[6] This technique was suggested by William Kruskal, to whom the authors wish to express their thanks.

TABLE 5

CONDENSATION OF TABLE 4

	On-Diagonal	Off-Diagonal	Total
High Pre-Wait status	7	4	11
Low Pre-Wait status	9	2	11
	16	6	22

Testing Hypothesis 4

Hypothesis 4 states that *Ss* seen by experienced therapists will improve more on the

TABLE 6

RELATION OF CHANGE IN THERAPY TO EXPERIENCE OF THERAPIST

(Change while in therapy)

	Q Adjustment		TAT Score	
	+	0−	+	0−
Exper.	6	6	8	4
Inexper.	7	3	1	9
		NS	$\chi^2 = 7.245$*	

* Significant at .01 level.

two test measures than those seen by inexperienced therapists. To determine if there were any bias in the selection of the cases seen by those two groups, the direction of change while waiting was examined. The experience of the therapist-to-be-seen was found to be unrelated to changes before therapy begins. Table 6 shows the relation of the direction of change while *in therapy* to the experience of the therapist. Before the very high degree of significance in Table 6 is taken as clear evidence of the positive effect of experienced therapists on the TAT changes of their patients, several other factors must be explored for their possible interaction.

1. Since length of In-therapy time showed a nonsignificant trend (.10) in the direction of longer cases changing negatively on the TAT, the first question of this kind is: are the experienced therapists seeing more cases that waited a shorter time to begin therapy and so had a shorter In-therapy period and a bias toward positive change? This was found to be a nonsignificant relationship in a chi square test.

2. Since Pre-therapy status on the TAT was found to be negatively related to In-therapy direction of change (see Table 3), are the experienced therapists perhaps seeing more cases with initially low scores? This, too, was found not to be the case using a chi square.

An on-diagonal, off-diagonal analysis confirmed these findings. The relation of the experience of the therapist to the direction of the TAT change was not significantly affected by the Pre-therapy status nor by the length of time in therapy. It seems clear that the experience of the therapist is the most powerful influence of those investigated here in producing positive change in the patient's TAT mental health rating during therapy.

Up to this point then, it can be said that there were two conditions under which the TAT mental health rating improved significantly: (*a*) without therapy, where the waiting period before beginning therapy was 8 to 24 weeks (Long-wait), and (*b*) with therapy, where the patient was seen by an experienced therapist. As far as the TAT is concerned, it seems a reasonable conclusion that over the time periods involved here, no therapist at all was better than an inexperienced one. Whether the gains made while waiting would have been sustained if no therapy had been entered into remains an open question, but a comparison of the mean changes under these

TABLE 7

MEAN SCORES FOR EXPERIENCED AND INEXPERIENCED GROUPS

	Pre-Wait	Pre-Therapy	In-Therapy	Post-Therapy	*t* Pre-In	*t* Pre-Post
Q Scores						
Exper.						
$N = 12$	39.2	39.7	41.0		—	
$N = 11$		39.0		46.0		2.536*
Inexper.						
$N = 10$	33.1	31.0	37.3		2.163*	
$N = 8$		32.81		40.25		1.563
TAT Scores						
Exper.						
$N = 12$	9.92	10.41	12.58		2.98**	
$N = 11$		10.27		12.72		2.190*
Inexper.						
$N = 10$	10.30	12.80	11.40		—	
$N = 8$		12.87		10.25		1.807

* Significant at .05.
** Significant at .02.

conditions confirms the above conclusion. The mean waiting period improvement for the Long-wait group was +2.9. The mean in-therapy time improvement for those seen by experienced therapists was +2.0. The *t* between these, .677, was not significant. The mean In-therapy change for those seen by inexperienced therapists was −1.4. The *t* between this and the improvement of the Long-wait group is 3.014, significant beyond the .01 level. Is the conclusion, then, that there is no difference between spontaneous remission and therapy change brought about by an experienced therapist and that inexperienced therapists have, if anything, a deleterious effect on their patients? As this conclusion would have strong implications, it is important that it be explored carefully. For this reason, the analysis was extended to the post-test data.

Table 7 reports the mean scores for both tests at the four test points for those seen by the two groups of therapists. A *t* test analysis for the differences between these means was used. The significant improvement between the Pre-therapy and the In-therapy test on the TAT of those in the experienced therapists' group has already been noted. A further result appeared to be a significant improvement on the Q adjustment score during this time of those in the *inexperienced* therapists' group. However, when an on-diagonal, off-diagonal analysis was used to examine the relationship of Q change in therapy and therapist experience for high and low Pre-therapy *Ss*, it was found that these proportions (8/11 on diagonal for low scorers) was significantly different from the proportion (4/11 off-diagonal for high scorers) at the .05 level. Thus, the improvement on Q adjustment for *Ss* in therapy with inexperienced therapists is largely a function of bias in Pre-therapy status. The *Ns* for the comparisons between Pre- and Post-therapy scores were reduced to 19 by the loss of 3 cases; 1 from the experienced and 2 from the inexperienced group. The results here show that patients seen by experienced therapists improve significantly by the end of therapy on both measures, whereas those seen by inexperienced therapists failed to improve significantly on the Q score and, in fact, came close to a significant deterioration on the TAT. As has been noted, those who waited a long time improved significantly on the TAT, but there was no commensurate shift of the Q adjustment score.

One difference between improvement with and without therapy has been found in this study. Those who waited a long time before beginning improved on the projective measure of their mental health but did not incorporate this into their self-picture as reflected by the Q score. Those who completed therapy with experienced therapists reflected their "deep" improvement on the conscious level in an improved self-description. Those completing therapy with inexperienced therapists not only did not improve their self-picture, but bordered on a deterioration in health.

Summary and Conclusions

Twenty-two self-referred psychoneurotic patients were tested at four points in time: (*a*) Pre-wait—on first being accepted as therapy cases and placed on the waiting list; (*b*) Pre-therapy—after a waiting time which varied in length; (*c*) In-therapy—after therapy had begun at a point equal to the waiting interval; and (*d*) Post-therapy. The instruments employed were the (*a*) Butler and Haigh Q sort, a self-descriptive test from which the Q adjustment score was computed, and (*b*) the TAT, a projective, from which a diagnostic rating was made. Four hypotheses were tested:

1. There will be more change during a period of therapy than an equal period of no therapy. This was confirmed for the Q adjustment score but not for the TAT.

2. Longer periods of therapy will be associated with more improvement than shorter periods. This hypothesis was not confirmed for either test.

3. Longer waiting periods will not be associated with more improvement than shorter periods. This was confirmed for the Q adjustment score but not for the TAT. Those who waited longer did in fact improve on the TAT, and those with shorter waits did not. The long wait improvers appear to be a spontaneous remission group with respect to the TAT.

4. The fourth hypothesis that patients in therapy with experienced therapists would improve more in the In-therapy period than

those in therapy with inexperienced therapists was confirmed for the TAT but not for the Q score. The fact that the positive effect of the experienced therapists was clearly independent of the Pre-therapy level of the TAT score and of the length of time in therapy made a strong case for the special effectiveness of these therapists. This in-therapy improvement could not be attributed to a continuation of a spontaneous remission trend.

The Post-therapy tests showed those in therapy with experienced therapists to have improved significantly on both tests, whereas those in therapy with inexperienced therapists not to have improved significantly on either and, in fact, they bordered on a significant decrease in health on the TAT.

It would appear that this study could conclude:

1. Therapy affects adjustment based on self-description, whereas equal waiting time does not.

2. A long waiting period for this type of patient improves the mental health picture as revealed by the TAT.

3. Equally long therapy does not necessarily improve the mental health picture.

4. An experienced therapist is effective in improving the mental health both as revealed by the TAT and as self-described.

5. An inexperienced therapist is ineffective in improving the self-picture and may actually be effective in decreasing the health as revealed by the TAT.

6. From the last two conclusions, it appears that therapists have special effects, depending on the level of their experience.

REFERENCES

BARRON, F., & LEARY, T. Changes in psychoneurotic patients with and without psychotherapy. *J. consult. Psychol.*, 1955, **19**, 239–245.

BUTLER, J., & HAIGH, G. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers & Rosalind Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 55–75.

CARTWRIGHT, D. S. Note on "Changes in psychoneurotic patients with and without psychotherapy." *J. consult. Psychol.*, 1956, **20**, 403–404.

DYMOND, ROSALIND. Adjustment changes over therapy from Thematic Apperception Test ratings. In C. R. Rogers & Rosalind Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 109–120. (a)

DYMOND, ROSALIND. Adjustment changes over therapy from self-sorts. In C. R. Rogers & Rosalind Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 76–84. (b)

DYMOND, ROSALIND. Adjustment changes in the absence of psychotherapy. *J. consult. Psychol.*, 1955, **19**, 103–107.

EYSENCK, H. The effects of psychotherapy: An evaluation. *J. consult. Psychol.*, 1952, **16**, 319–324.

(Received March 17, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

TOWARD THE ELIMINATION OF THE CONCEPT OF NORMALITY [1]

DAVID FREIDES
Lafayette Clinic, Detroit, Michigan

In recent years, the literature concerning the nature of normality in human personality has shown increasing support for an "idealist" conception of psychological normality. Normality is viewed as an approximation to an ideal or combination of ideals, such as health, rationality, honesty, integration, maturity, or morality. This literature has justifiably criticized the traditional approaches to this problem, i.e., a "symptom-free" conception, specified as absence of pathological stigmata, or a "centrist" conception, specified either as the statistical average of normally distributed characteristics or as the balance between tendencies, which, in their extremes, are considered pathological, like introversion and extroversion. This paper will present a critique of normative perspectives in personality theory, focusing on the "idealist" position (criticism of traditional approaches may be found in works by Hartman, 1939; Maslow, 1954; Mowrer, 1954; Shoben, 1957, and others), and propose an alternative to these approaches.

A common starting point among authors proposing an idealist conception of normality is the recognition of the enormous extrabiological potential for variation inherent in human beings as evidenced by complex cultural activity, symbolization, and altruism. They stress that these inherently human qualities deserve more serious attention than they are getting from students of personality and should be embodied in the concept of normality. Shoben (1957, p. 185), for example, states, "the fundamental contention advanced here is that behavior is 'positive' or 'integrative' to the extent that it reflects the unique attributes of the human animal." As the "unique attributes" of humans are extremely varied and frequently repugnant, the writers go on to specify what they mean. Maslow (1954, Ch. 12) lists 17 characteristics of self-actualizing people including acceptance of self, others, and nature; spontaneity; problem centering; the quality of detachment; autonomy; democratic character structure; creativeness; and resistance to enculturation. Bond (1952) mentions freedom to focus energy on main purposes, ability to work and love with ease, and to achieve happiness and efficiency somewhat in proportion to circumstances. Shoben (1957) writes of self-control, personal responsibility, social responsibility, democratic social interests, and ideals. McLaughlin (1950) enumerates emotional independence and self-reliance; balance between giving and getting; relative freedom from egotism, inferiority feelings, and excessive competitiveness; conscience; genital sexuality; constructive aggressiveness; solid sense of reality; flexibility; and adaptability.

Several major difficulties and limitations are inherent in this approach. First, it establishes as absolute desiderata patterns of behavior and values that grow out of a particular cultural context during a particular period of its history. It assumes that what is good for us is good for everybody everywhere. Several authors explicitly recognize that these criteria of normality involve assumptions in the nature of value judgments. Other writers do not, but cultural anthropology and history provide evidence that freedom from inferiority feelings or genital sexuality or constructive aggressiveness, for example, are not and have not been universally considered desirable goals for human behavior.

[1] Part of this paper was prepared while the writer was at the Army Medical Service School, Ft. Sam Houston, Texas.

Second, this approach generally does not take into account the circumstances under which the personality is to function. It implies that variations in circumstances do not determine or even influence the functioning of the personality and precludes the study of how behavior and circumstance covary. Third, it largely assumes that postulated criteria of normality are positively correlated, or can be positively correlated. As to whether they are correlated, reference can be made to several attempts to study "normal, healthy individuals." Maslow (1954) and Bond (1952), for example, went to great efforts to obtain suitable examples. Bond guarded against his own selection biases by studying almost complete populations of student council members and concluded nonetheless that 57% could benefit from ideal mental health services. Maslow grudgingly devotes two pages to the imperfections of self-actualizing people. Roe's (1950) studies of eminent scientists (to be sure, using criteria of success and not of normality) amply indicate that nonnormal characteristics regularly accompany achievement, creativity, and success. Redlich (1952) reviews several studies of "normal" populations in which "rich pathological" material was found.

Perhaps even more significant is the question of whether it is reasonable to anticipate that characteristics of normality such as those under discussion can be positively correlated. Consider, for example, such positively valued attributes as love and independence. Ideal patterns of loving devotion and enduring commitment to one's beloved, coupled with independence, assertiveness, and freedom from infantile demands of succorance, are generally accepted as positive characteristics of the normal; but the conflictual nature of multiple human needs and aspirations is obscured by an approach which does not recognize that love and commitment *necessarily* and inevitably restrict independence.

Finally, a value based conception of normality tends arbitrarily to slant research so as to preclude investigations of the relationships between devalued patterns of behavior and ideal patterns of behavior.

These criticisms are not to be construed as arguments for banning values from consideration in personality theory. The question is where and how to deal with them in a way that is scientifically appropriate. Values and value systems may be studied objectively as variables, and the task of psychology is to assess their relationships with other variables and with behavior. Such an approach does not imply that every moral and ethical system is as good or as workable as any other nor that the search for the "good" in either individual or culture should be minimized, but it does indicate a way in which such problems may be considered at levels of discourse which are germane to science. In this manner science is not faced with questions it cannot answer.

It is significant that the proponents of the point of view under discussion seem to be all for the good life but are rather vague as to what they mean by it, while theologists and ethicians are quite explicit and vigorous defenders of the positions they take. The apparent proclivity for tolerance, understanding, and minimization of differences, characteristic of this school of thought (with few exceptions), conceals and overrides hard differences in the approach to life embodied in different value systems.[2] Yet, to consider values as variables seems definitely uncongenial to those who hold that the values they profess are the only values worthy of consideration and study. For example, Shoben (1957), while arguing that a normative approach based on a statistical average precludes the criticism of a culture, confuses a conflict of values with a scientific controversy when he disdains the possibility that the storm trooper would be considered the prototype of integrative adjustment in the Nazi culture. Disapproval of Nazi culture has nothing to do with understanding how a man raised in Nazi Germany might be expected to become a storm trooper. Science can do no more than predict the consequences of a value system. Political, religious, or cultural conflicts are fundamentally not resolved in the laboratory or on the couch.

Writers adopting an idealist approach con-

[2] See Herberg (1957) for a critical discussion of the ethical inadequacies of theoretical orientations in psychology which minimize differences in value alternatives and preclude the conceptualization of the realities of moral conflict in the individual.

cerning psychological normality take a position that is steeped in contradiction. On the one hand, their conception of normality has an absolutist quality that is laced with culture-bound value assumptions. On the other hand, they show marked reservations concerning a conception of normality based on social conformity and they abhor any taint of absolutism. Maslow (1954), for example, specifies autonomy, independence of culture and environment, and resistance to enculturation as characteristics of the self-actualizing (normal) person. In doing so, Maslow evidently means to divorce himself from purely statistical conceptions of social conformity. Certainly he does not mean compulsive autonomy, or compulsive resistance to enculturation, yet his whole perspective, shared with many others, including Fromm (1941), Mowrer (1954), and Shoben (1957), is one of meeting a specified ideal, and this inevitably involves some form of cultural bias, some form of limitation, some constriction of freedom.

In a discussion of this contradiction, Knight (1946) deals with the problem by (*a*) specifying that freedom is largely a subjective matter; (*b*) showing that many subjective feelings of freedom are spurious, and (*c*) concluding that genuine freedom and mental health involve the sense of inner compulsion and conformity based on a rational appreciation by the individual of the assumptions, conditions, motives, and values in his life. It is this bedrock of rationality to which all of these writers ultimately arrive.

Kubie (1954) takes perhaps the extreme position and asserts that the sole criterion for normality is the predominance of conscious and preconscious over unconscious factors in the determination of an act. Since Kubie rules out consciousness, per se, as a criterion (a delusion is conscious but not normal), he can only be interpreted to mean that rational, reasonable, reality-oriented behavior is normal. On the surface, this seems to be a satisfactory solution to the problem. It posits a criterion which is neutral as far as value is concerned and can conceivably lead to objective assessment procedures. However, such a point of view seemingly would label all nonrational behavior and motivation as abnormal and in doing so would devalue and implicitly condemn any manifestation of emotional spontaneity. Arguing, as Kubie does, that it is the origin of the act and not its subsequent automatic execution that must be assessed, would nonetheless require, at some point, meticulous evaluation of matters of taste, preferences in art, religious values and beliefs, and love objects in order to fulfill this criterion. Kubie's disclaimer (1954, p. 187), "Here is no unreal fantasy of a 'normal' individual out of whom all the salty seasoning of secondary unconscious motivations has been dissolved," when followed in the next paragraph by "What was unconscious . . . must become accessible enough to self-inspection to become conscious when needed," suggests that his conception of "salty seasoning" is not so spicy after all.

When the search for the normal leads to the elevation of the rational, a host of problems concerning the nature of reason, logic, and rationality may be overlooked. These include the distinctions between a rational process and rational content (Hartman, 1956; Reider, 1950), the illogical and prelogical factors which contribute to the development of intelligence and reason (Piaget, 1953), the role of unconscious and regressive factors in creativity (Kris, 1952), and the process of change in form and content of knowledge and logic through a history which has not ended and whose future course is unpredictable by any reliable means.

Finally, and this criticism is germane to practically every attempt to define normality, it assumes that rational, conscious, and preconscious processes, once established, carry on of their own accord without environmental support and stimulation. Recent studies on the effects of sensory deprivation (Hebb, 1955), the classic studies of perception under distorted conditions (Lawrence, 1949), and many studies of behavior under stress (Lazarus, Deese, & Osler, 1952) indicate that this is simply not the case. To specify consciousness or rationality as the essence of normality merely begs the question.

A truly objective, scientific solution to the problem of defining normality when one is concerned with something more than averaging existing traits and behaviors, when one seeks to embody in this definition a particu-

lar conception of what man may be, is impossible to attain. As soon as values, symbols, altruism, and the like are specified in the definition of normality, objectivity yields to time-bound and culture-bound assumptions which are not subject to scientific verification and which are of limited generality. That each person deliberately or unconsciously has a scheme of values, beliefs, and ethics which imparts to him his uniquely human character, that, indeed, to be a human requires such a scheme, is a proposition that does not establish a *particular* system of values or logic as an absolute criterion of normality.

Psychology should stop asking the question "What is normal?" Personality theory requires another orientation. It seems to me that, for purposes of scientific theory construction and also for practical clinical purposes, a rather different and potentially more fruitful approach can be taken. This would entail the viewpoint that human beings have a variety of potentialities, and that the achievement of certain potentialities may entail certain limitations and that achievement and limitation vary with conditions.

Consider it is the distant, withdrawn personality who is unable to form close relationships who, it has been said, stands the best chance of enduring the stress of a prolonged flight into space. The suspicious mistrusting individual, difficult to get along with, may, because of these very characteristics, be an excellent tax inspector, research scientist, or counterintelligence agent. The vain narcissistic personality can become a great entertainer. Reverse the point. Is great acting possible without narcissism? Is creative research possible without mistrust? Is prolonged isolation tolerable for a gregarious individual? The legitimate task of science is not to stand back smugly and label the various distributions of energy and effort as abnormal or normal. The task of theoretical psychology is to provide the concepts and principles which would enable us to comprehend achievements and failings *and their interrelationships*.

Such an approach focuses on man's behavior as the data available for scientific study. Existing terms and concepts such as repression, motive, conformity, defense mechanism, and habit need not be discarded as conceptual tools, but with this perspective the scientist does not imply damnation or praise when he uses them to characterize a psychological process. This approach sets no limits on the sophistication of the theories developed to encompass the facts of behavior, but it lessens the likelihood of confounding facts and theories and value assumptions. It avoids positing criterion value systems but freely admits the incorporation of value systems as variables contributing to or limiting different potentialities. It sets the stage for research into the development and exercise of man's potentialities for what, on other grounds, may be considered good or evil, but does not prejudge or condemn. It permits the description and understanding of man as he was, is, and can be.

Several features of the present position have been elaborated in the literature. Marmor and Pumpian-Mindlin (1950) and Jahoda (1953), for example, have stressed the interaction of personality and situational variables in the determination of behavior. Hartman (1951) points out that conflicts are inherent to human existence and (1939) that rationality and freedom do not necessarily imply health, while regression and defense are not necessarily maladaptive. Schafer (1954, p. 172) writes, "The growth and organization of personality appears to require that certain of each individual's potentialities be cultivated and others neglected or even plowed under." Marzolf (1947) concludes, "In fact there is no need for the term 'mental disease.'" Several writers (Darrah, 1939; Jones, 1942; Wile, 1940) are doubtful about the existence of "normal" personalities. But in one way or another, they all (with the exception of Wile, 1940) cling to the traditional perspective and express or reflect some concept of normality.

If the framework proposed is theoretically acceptable, objections may nevertheless be raised on practical grounds. If no valid conception of normality is possible, is treatment ever indicated and, if so, toward what goals is treatment directed? At present, certain personality patterns can be labeled abnormal and treatment is presumably directed toward changing abnormal personalities or patterns of behavior into normal ones. The definition of what is normal is considered critical to

clinical practice and its elimination apparently leaves an enormous conceptual void.

Actually, there are several unwarranted assumptions implicit in the view that a conception of normality is central to clinical practice. First is the assumption that treatment is initiated with the finding of abnormality, however defined. Operationally, treatment is initiated when motivation from some source exists for treatment and when therapists are motivated and prepared to undertake treatment. There is a reciprocal relationship between professional personnel and patient pool in that the professional proffers the expectation of relief from distress and contributes to the development of this anticipation in potential clients. Absence of some relevant form of this expectation seems to be one of the factors underlying failure in the treatment of lower-class patients by predominantly middle and upper-class professionals (Hollingshead & Redlich, 1958). The point is that the initiation of treatment results less from a finding of abnormality than from the meshing of reciprocally intertwined (more or less) motivations.

Second, the notion that the aim of treatment consists of changing the patient from abnormal to normal is unjustified. In many cases, in those perhaps most amenable to current forms of treatment, therapy consists largely of helping the patient to resolve conflicting trends within his personality rather than making him or helping him to become normal. The patient who has successfully completed this type of treatment (in which the therapist can remain relatively neutral with regard to values) does not by any flight of the imagination meet the usual criteria of normality. (The individual who speaks of his having been analyzed or treated psychotherapeutically and implies that he is "normal" is usually labeled a prig.) Furthermore, as already noted, if the patient has been enabled to deal more successfully with certain kinds of problems and conflicts, this may be at the expense of other potentialities.

A third questionable point is the assumption that the goals of treatment are determined by scientific criteria. In many types of cases the resolution of conflict is only a minor part of therapy. Much more than the bolstering or the elimination of defensive patterns within the personality seems indicated in certain types of schizophrenia, in severely schizoid personalities or in antisocial personalities. Very many of these cases require that a way of life be taught. In other words, certain cases require far more than reorganization of the personality; they require education itself and the inculcation of values. Once again, it seems to me that science can contribute to an understanding of the educative process but that the content of this education is largely a matter that must be decided on other grounds. Here, the articulation and specification of the values under which or toward which the therapist operates is of great importance, ***but these choices are not scientific matters.*** We do the best we can much like any parent living at a particular time and in a particular place.

This point, that values, symbols, and the like *are* involved in psychotherapy (Hacker, 1945), brings the discussion full circle. Davis (1938, p. 65) in a little cited analysis of the mental hygiene movement has argued that the social function of the mental hygienist "is not that of a scientist but that of a practising moralist in a scientific, mobile world." It seems to me that what Davis is saying is that professionals dealing with "ways of life" inevitably have to rely on *some* consistent framework to fill gaps and resolve conflicts where they exist and that this framework willy-nilly will be that prevalent in the cultural group to which the professionals belong. If this issue may be obscured in the treatment of adults, it is blatantly apparent and quite inevitable in the treatment of children and in the advice given to parents about techniques of child rearing (Erikson, 1950). As a consequence, it seems likely that psychological treatment is a conservative force in any society no matter how radical the theories about psychotherapy sometimes appear to be. But the implicit or explicit commitment to values involved in the practice of psychotherapy is an issue separate and distinct from a scientifically useful specification of the concept of normality. The future development of a scientific understanding of personality depends on a less restrictive perspective which, ironically, should benefit clinical practice as well.

A shift in perspective in personality theory from one which looks upon man as normal or

abnormal or somewhere in between to one which views him as having varying potentialities and limitations under varying conditions, seems to offer an orientation more viable, secure, and fruitful than that prevailing at present. The viable aspect of this position comes out of its capacity to conceive of man in different times and different circumstances. The security features result from the elimination of the necessity for the clinician to rationalize about his own normality (Reider, 1950), permitting his work to be evaluated objectively in terms of its results. Finally, the fruitfulness of this approach rests on its reluctance to condemn, to label abnormal, and its emphasis on the potentialities of every person under the proper conditions. It sets the stage for more extensive investigations into the nature of these potentialities and the conditions of their achievement, all the while keeping before us the pitfalls and limitations that such achievement may involve.

REFERENCES

Bond, E. D. The student council study: An approach to the normal. *Amer. J. Psychiat.*, 1952, **109**, 11–16.

Darrah, L. W. The difficulties of being "normal." *J. nerv. ment. Dis.*, 1939, **90**, 730–739.

Davis, K. Mental hygiene and the class structure. *Psychiatry*, 1938, **1**, 55–65.

Erikson, E. H. *Childhood and society*. New York: Norton, 1950.

Fromm, E. *Escape from freedom*. New York: Rinehart, 1941.

Hacker, F. J. The concept of normality and its practical significance. *Amer. J. Orthopsychiat.*, 1945, **15**, 47–64.

Hartman, H. Psychoanalysis and the concept of health. *Int. J. Psycho-Anal.*, 1939, **20**, 308–321.

Hartman, H. Ego psychology and the problem of adaptation. In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951.

Hartman, H. Notes on the reality principle. *Psychoanal. Stud. Child*, 1956, **11**, 31–53.

Hebb, D. O. The mammal and his environment. *Amer. J. Psychiat.*, 1955, **111**, 826–831.

Herberg, W. Freud, religion and social reality. *Commentary*, 1957, **23**, 277–284.

Hollingshead, A. B., & Redlich, F. C. *Social class and mental illness*. New York: Wiley, 1958.

Jahoda, Marie. The meaning of psychological health. *Soc. Casewk.*, 1953, **34**, 349–354.

Jones, E. The concept of the normal mind. *Int. J. Psycho-Anal.*, 1942, **23**, 1–8.

Knight, R. P. Determinism, "freedom" and psychotherapy. *Psychiatry*, 1946, **9**, 251–262.

Kris, E. *Psychoanalytic explorations in art*. New York: International Univer. Press, 1952.

Kubie, S. The fundamental nature of the distinction between normality and neurosis. *Psychoanal. Quart.*, 1954, **23**, 167–204.

Lawrence, M. *Studies in human behavior*. Princeton: Princeton Univer. Press, 1949.

Lazarus, R. S., Deese, J., & Osler, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.

McLaughlin, J. T. Normality and psychosomatic illness. *Ment. Hyg., N. Y.*, 1950, **34**, 19–33.

Marmor, J., & Pumpian-Mindlin, E. Toward an integrative conception of mental disorder. *J. nerv. ment. Dis.*, 1950, **111**, 19–29.

Marzolf, S. S. The disease concept in psychology. *Psychol. Rev.*, 1947, **54**, 211–221.

Maslow, A. H. *Motivation and personality*. New York: Harper, 1954.

Mowrer, O. H. What is normal behavior? In L. A. Pennington & I. A. Berg (Eds.), *An introduction to clinical psychology*. (2nd ed.) New York: Ronald, 1954. Pp. 58–88.

Piaget, J. *The origin of intelligence in the child*. London: Routledge & Kegan Paul, 1953.

Redlich, F. C. The concept of normality. *Amer. J. Psychother.*, 1952, **6**, 551–569.

Reider, N. The concept of normality. *Psychoanal. Quart.*, 1950, **19**, 43–51.

Roe, Anne. The use of clinical diagnostic techniques in research with normals. In M. Reymert (Ed.), *Feelings and emotions*. New York: McGraw-Hill, 1950. Pp. 336–342.

Schafer, R. *Psychoanalytic interpretation in Rorschach testing*. New York: Grune & Stratton, 1954.

Shoben, E. J. Toward a concept of the normal personality. *Amer. Psychologist*, 1957, **12**, 183–189.

Wile, I. S. What constitutes abnormality? *Amer J. Orthopsychiat.*, 1940, **10**, 216–228.

(Received March 31, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

AGE DIFFERENCES IN THE INTERSCALE FACTOR STRUCTURE OF THE GUILFORD–ZIMMERMAN TEMPERAMENT SURVEY

A. W. BENDIG
University of Pittsburgh

One of the relatively unexplored areas in objective personality measurement concerns age differences in test scores. Many manuals for personality tests state that an age correction must be made in the raw scores before the scores can be referred to the test norms, or norms are given for different age groups without any apparent interest in what these age differences mean in terms of possible changes in personality integration. A second question relates to possible age differences in the internal structure or interrelationships among the scales included in the personality inventory. Most of these inventories have been developed, by factor analysis or some other statistical technique, on relatively young groups of subjects (the traditional college sophomore being widely used), and the scale scores are subsequently interpreted under the assumption that each scale means the same thing or measures the same underlying personality dimension at all age levels. This appears to be a hazardous assumption in the absence of data on the stability of personality test scores over the mature age range.

A personality inventory that is widely used with nonpsychiatric populations is the Guilford-Zimmerman Temperament Survey (Guilford & Zimmerman, 1949) which contains 10 scales of General Activity (G), Restraint (R), Ascendance (A), Sociability (S), Emotional Stability (E), Objectivity (O), Friendliness (F), Thoughtfulness (T), Personal Relations (P), and Masculinity (M). These scales are revised and condensed versions of 13 scales that had been developed by factor analyses of test items and published in previous inventories (Guilford, 1940; Guilford & Martin, 1943a, 1943b). Although the scales presumably measure 10 orthogonal or independent first-order personality dimensions, scores from certain scales typically show rather large intercorrelations. For example, correlations between the A and S scales usually are about .68 and between the E and O scales average around .66. However, the correlations among other scales are quite low. This suggests that scores from these scales may be organized at a higher level of personality into a smaller number of more general factors describing broader syndromes of behavior. Factor analyses of the correlations among the GZTS scales would provide evidence as to the number of such factors involved and an empirical method of describing these syndromes. However, as pointed out by Guilford and Zimmerman (1956, p. 2), these second-order or "super" factors represent factors common to the inventory *scores* and do not necessarily represent the second-order organization among the original first-order inventory *factors* which the scores were designed to measure.

The research reported below was designed to provide evidence as to the factor organization of the GZTS scores and to investigate possible age differences in both the average scale scores and the factor loadings of the individual scales.

PROCEDURE

Subjects. The *Ss* were 400 men whose GZTS scores were available from the files of a local psychological service agency.[1] Over 800 male and female *Ss* were tested during the year 1953, and the cases were randomly selected from this larger group with the restrictions on randomness being: (*a*) only male *Ss* were included; (*b*) 100 *Ss* were selected from each of the following age (at time of testing) ranges—20–29, 30–39, 40–49, and 50–59 years; and (*c*) the

[1] The author is indebted to James Horgan and the research staff of Psychological Service of Pittsburgh for their kindness in making these data available.

TABLE 1

SUMMARY OF ANALYSES OF VARIANCE OF AGE DIFFERENCES IN MEAN SCORES OF MEN ($N = 400$) ON THE GUILFORD-ZIMMERMAN TEMPERAMENT SURVEY

GZTS Scale	Age Groups								Error Mean Sq. ($df = 396$)	Significance of Age Differences (F)			
	20–29		30–39		40–49		50–59			Groups ($df = 3$)	Linear ($df = 1$)	Quadratic ($df = 1$)	Cubic ($df = 1$)
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*					
G	18.86	4.69	18.87	5.06	18.85	5.17	17.38	5.16	25.49	2.15	3.90*	2.15	.40
R	18.87	4.21	20.18	4.32	20.64	3.52	20.66	4.00	16.36	4.33**	10.39**	2.54	.05
A	19.83	4.97	18.86	5.00	17.61	5.63	16.74	5.65	28.62	6.47**	19.33**	.01	.76
S	23.77	5.01	22.88	4.76	21.84	5.12	20.60	6.22	28.45	6.56**	19.56**	.11	.00
E	21.87	4.75	21.57	5.25	20.96	4.90	21.58	4.27	23.33	.63	.47	.91	.51
O	21.87	4.75	21.57	5.25	20.96	4.90	21.58	4.27	23.33	1.87	3.32	.60	1.69
F	17.39	4.73	18.42	4.81	17.45	4.75	17.99	4.71	22.79	1.04	.15	.26	2.70
T	19.63	4.28	19.67	4.56	20.53	3.74	20.31	4.01	17.47	1.18	2.41	.10	1.03
P	21.61	5.38	23.53	3.90	22.11	4.41	22.77	4.10	20.31	3.41*	1.04	1.95	7.23**
M	20.96	3.20	20.14	3.55	19.54	3.25	19.14	3.51	11.55	5.43**	15.90**	.38	.00

* $P < .05$.
** $P < .01$.

age distribution within each age decade was kept as rectangular as possible. The mean ages for the four groups of *Ss* were 24.8, 34.5, 44.2, and 54.0, while the standard deviations of ages within each group were 4.4, 2.9, 3.0, and 2.5. The *Ss* had been referred to the agency by local business firms for psychological testing for employment, promotion, or company personnel survey purposes and had been given a large battery of aptitude, interest, and temperament tests including the GZTS.

Method. The raw scores from each of the 10 GZTS scales were subjected to a one-criterion (between-within) analysis of variance with the age of the *Ss* (four decades) being the independent variable. The between-age groups sums of squares were further analyzed into linear, quadratic, and cubic components by the method of orthogonal polyomials (Grant, 1956) and tested for significance. The F ratio test of the between-groups sums of squares tests for the presence of significant differences among the GZTS score means of the age groups, but does not specify the form of the functional relationship between age and the scale means. The further breakdown of the sums of squares tests whether the relationship is a straight line (linear), simple curvilinear (quadratic), or more complex (cubic) change in means with increasing age.

The scores within each age group were intercorrelated by the usual product-moment method, and each of the four matrices of correlations was factored by the complete centroid method, with five factors being extracted from each matrix. Each successive factor was tested for significance using Humphrey's rule (Fruchter, 1954, pp. 79–80), and the apparently significant factors were rotated to orthogonal simple structure using the normalized varimax criterion (Kaiser, 1958).[2]

[2] The author's appreciation is extended to Gary Lotto and William B. Kehl of the University of Pittsburgh Computation and Data Processing Center for making their facilities available for this phase of the statistical analysis.

RESULTS

A summary of the analyses of variance of the GZTS scales can be found in Table 1, which gives the means and standard deviations of the age groups on each scale, the within-age groups (error) mean square from the analyses of variance, and the F ratio tests of significance of the between-age groups mean square and its linear, quadratic, and cubic components. The age differences in mean scores were significant at the .01 level for four scales (R, A, S, and M), and one additional scale (P) was significant at the .05 level. The linear trend in the means was significant at the .01 level for four scales (R, A, S, and M) and at the .05 level for the G scale. The cubic trend was significant at the .01 level for the P scale. None of the quadratic components was significant even at the .10 level. The means of the A, S, and M scales show a linear (straight line) decline with increasing age, while the G scale is essentially level until age 50 and shows a sharp drop between the 40–49 and 50–59 age groups. The R scale mean scores increase with age with some deceleration in the rate of increase. The P scale means tend to increase with age, but the increase is irregular, with the largest P scale mean found for the 30–39 age group. Scales E, O, F, and T show no systematic variation with age.

The first two centroid factors were significant at the .01 or .05 levels within all four age groups, while the third factor was sig-

nificant at the .01 level for the 20–29 and 30–39 groups but was not significant for the two older groups of *Ss*. Neither the fourth nor the fifth factor approached significance at the .05 level within any age group. For comparability between age groups, the first three factors were retained in the subsequent normalized varimax rotations with the remaining two factors being dropped. The percentages of variance accounted for by the first three orthogonal factors within each group were 54.1, 51.3, 45.1, and 44.3.

The rotated factor loadings and communalities for each GZTS scale can be found in Table 2. The approximation to simple structure was moderately successful with slightly over half of the loadings having absolute values less than .20. The first factor is characterized by consistently high loadings on the O, P, and F scales for all age groups and high to moderate loadings on the E and M scales only for the younger groups. The second factor shows consistently high loadings on Scales S and A and moderate loadings on Scale G for all age groups and gradually increasing loadings on Scale E as the age of the *Ss* increases. The third factor loads Scales R and T only with no consistent loadings on any other scale. The O scale has a small but consistent loading on the second factor, and Scales T and R have slight negative (T) and positive (R) loadings on the first factor. The similarity between comparable factors appearing in the four analyses was quantified by the method suggested by Wrigley and Neuhaus (1955). The indices of factor similarity for the first rotated factor ranged from .98 to .83, with the median of the six indices being .93. Indices for the second factor ranged from .96 to .68 with the median also being .93, while the range for the third factor was .96 to .41 with a median of .85. The indices were consistently lowest when factors were compared between the two groups at the extremes of the age continuum, 20–29 and 50–59, and is probably attributable to the changing factor composition of Scales E and M.[3]

Discussion

The first factor found in the factor analyses of the GZTS scales included the scales of Objectivity, Friendliness, and Personal Relations. These three scales are GZTS versions of the Objectivity, Agreeableness, and Cooperativeness scales developed by Guilford and Martin (1943b) from a factor analysis of items

[3] Tables giving the original correlation matrices, unrotated factor loadings, transformation matrices, tests of the significance of the factors, and indices of factor similarity have been deposited with the American Documentation Institute. Order Document No. 6120, remitting $1.25 for microfilm or $1.25 for photocopies.

TABLE 2

Rotated Orthogonal Factor Loadings (Decimal Points Omitted) of the Guilford-Zimmerman Temperament Survey Scales for Four Age Groups of Men

(N = 100)

GZTS Scale	Age 20–29				Age 30–39				Age 40–49				Age 50–59			
	Fr	SA	EI	h^2	Fr	SA	EI	h^2	Fr	SA	EI	h^2	Fr	SA	EI	h^2
G	−03	54	06	30	11	55	−20	36	−09	38	−13	17	−07	43	00	19
R	22	03	68	52	11	−07	65	44	−07	−11	37	15	31	−03	52	36
A	16	77	20	67	13	84	09	73	20	82	16	75	12	73	−07	55
S	10	85	−11	74	11	73	00	54	10	85	08	74	17	69	−32	60
E	77	32	−06	70	70	48	29	81	37	46	−32	45	35	70	15	64
O	90	19	−02	84	78	34	−01	72	73	19	−41	75	72	30	−04	62
F	57	−12	13	36	64	−14	34	55	55	−34	−31	51	77	−02	15	61
T	−18	09	69	51	34	05	49	35	−12	19	56	36	−06	−04	44	20
P	62	26	−04	46	51	20	−12	31	74	02	−05	55	74	08	03	55
M	53	−13	02	30	53	10	−16	32	34	07	−02	12	22	22	−14	12
Percentage of variance	25.1	18.5	10.2	53.8	22.2	19.6	9.5	51.3	17.1	19.6	8.7	45.4	19.8	18.3	6.4	44.5

reflecting "paranoid disposition." Thurstone (1951) reanalyzed the intercorrelations between 13 Guilford scales (including Objectivity, Agreeableness, and Cooperativeness) reported by Lovell (1945) and found nine second-order factors, one of which (Friendliness) had loadings only on the three Guilford-Martin scales. Guilford and Zimmerman (1956) reanalyzed Lovell's basic data, separating the items on the 13 scales into 70 short subscales, and factor analyzed the matrix of intercorrelations among the 70 variables. Three of the rotated orthogonal factors could be clearly identified as Objectivity, Agreeableness, and Cooperativeness. Although these factors were orthogonal, of the 26 subscales that were significantly loaded for at least one of these factors, 35% also had significant loadings on at least one of the other factors in this trio. Klugh and Bendig (1959) correlated the 10 GZTS scales with the ACE test and found that the O, F, and P scales were the only scales significantly correlated with the measure of general verbal intelligence. The evidence strongly indicates that the O, F, and P scales consistently define a personality factor that may be called Friendliness (Fr).

The second GZTS factor was defined by the Sociability, Ascendance, and (to a lesser extent) General Activity scales. Thurstone (1951) also found a second order factor loaded only on the S and A scales (Social Leadership) while Guilford and Zimmerman (1956) found four subscales loaded on both S and A factors. Again S and A appear to define a stable second-order GZTS factor which also is linked for the younger *Ss* with the G scale. Because of this linkage with a G and inadequate evidence that Leadership is reflected in the factor, it appears more parsimonious to call this second order GZTS factor Social Activity (SA).

GZTS Restraint and Thoughtfulness scales consistently defined the third factor. It is quite obvious that this is the Reflective factor found by Thurstone (1951) and is the reverse of the second-order Extraversion factor discussed at length by Eysenck (1953, pp. 103–128). To distinguish this GZTS factor from Guilford's first-order Rhathymia factor, we will use the name Extraversion–Introversion (EI).

The Emotional Stability scale appears to have loadings on all three second-order factors. It is probable that the E scale is measuring a second-order factor rather than a first-order factor, since Guilford and Zimmerman (1949) combined first-order factor scales of Depression and Cycloid Disposition in developing the GZTS E scale. Since there is only one scale on the GZTS that measures this second-order factor, a separate (fourth) factor of Emotionality (Em) could not appear in the analyses. The E scale divided its common factor variance primarily between the Fr and SA factors with considerable intersample fluctuation in the size of its loadings on these second-order factors.

The evidence seems rather strong that scores from the 10 GZTS scales can be organized into at least three, and possibly four, independent (orthogonal) factors: Friendliness (Fr), Social Activity (SA), Extraversion–Introversion (EI), and Emotionality (Em). The relative independence of the first three factors is confirmed by the correlations reported by Thurstone (1951) between his second-order oblique factors of Friendliness, Social Leadership, and Reflectiveness. Although rotation of oblique factors permits large intercorrelations among the factors (while orthogonal rotation permits only zero correlations among factors), Thurstone found the correlations to be .01, −.11, and .07. However, the results with the GZTS E scale suggest that the hypothetical Em second-order factor may be correlated with the Fr and SA factors.

Age differences are evident on two of the three major factors. Of the three scales defining the Fr factor, mean scores of O and F remained relatively constant between the age groups, and while the P scale scores tended to increase with age, they increased irregularly, reaching their peak in the 30–39 years age group. The first-order factor of Personal Relations or Cooperativeness may vary with age, but the broader factor of Friendliness remains relatively stable. Masculinity (GZTS Scale M) tends to be associated with Fr among the younger *Ss*, but as the mean scale M scores decline with age (contrary to the stability of the other scales defining Fr), the loading of this scale on Fr decreases to insignificance. Among the more generally mas-

culine groups of younger male *Ss*, individual differences in Masculinity are positively associated with differences in Friendliness, while within the generally more feminine groups of older *Ss* the correlation between Masculinity and Friendliness disappears.

All three of the scales loading on the factor of Social Activity show decreasing mean scores with increasing age. The loading of the GZTS G (General Activity) scale on factor SA slightly decreases with age with the size of the loading tending to parallel the size of the G scale mean. The older male *Ss* appear on the GZTS to withdraw more from vigorous social activities and to be more submissive in the social world. This does not imply a withdrawal from society but, rather, a reorientation of their role in society toward a less competitive position that is more free of the demands of social leadership.

The factor of Extraversion–Introversion shows some changes with age, primarily in the direction of greater introversion among older *Ss*. This trend was statistically significant only for the Restraint component of EI, with Thoughtfulness not increasing as markedly among the older groups. Scores on the two scales loaded for the EI factor increase up to the 40–49 age group, and no change in mean R and T scores is apparent between the 40–49 and 50–59 age groups.

Obviously this cross sectional study does not imply that male *Ss* show personality *changes* with age that are discussed above, but that these personality *differences* existed among groups of *Ss* who were born in the years (approximately) 1928, 1918, 1909, and 1899. Whether the trait differences apparent in 1953 reflect age changes in the *Ss* or differences in the environmental factors influencing personality formation in childhood during these early years cannot be assessed from this type of study.

Summary

Scores on the 10 Guilford-Zimmerman Temperament Survey scales were obtained for four groups of male *Ss* ($N = 100$) in different age decades: 20–29, 30–39, 40–49, and 50–59 years. Scores on the General Activity, Ascendance, Sociability, and Masculinity scales significantly and linearly decreased with increasing age, while the Restraint scores linearly increased with age and scores from the Personal Relations scale showed an irregular increase. Factor analyses of the scale intercorrelations resulted in three major orthogonal factors: Friendliness (Fr), Social Activity (SA), and Extraversion–Introversion (EI). Scales of Objectivity, Friendliness, and Personal Relations consistently loaded on the Fr "super" factor, Ascendance and Sociability loaded on the SA factor, while Thoughtfulness and Restraint loaded for the EI factor. The loading of Masculinity on the Fr factor showed a progressive decline with increasing age. Emotional Stability inconsistently loaded on both Fr and SA factors and it was suggested that the E scale may represent a fourth second-order factor (Emotionality) present in the GZTS inventory.

References

Eysenck, H. J. *The structure of human personality*. London: Methuen, 1953.

Fruchter, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.

Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Guilford, J. P. *An inventory of factors STDCR*. Beverly Hills, Calif.: Sheridan, 1940.

Guilford, J. P., & Martin, H. G. *An inventory of factors GAMIN*. Beverly Hills, Calif.: Sheridan, 1943. (a)

Guilford, J. P., & Martin, H. G. *Personal inventory: Manual of directions and norms*. Beverly Hills, Calif.: Sheridan, 1943. (b)

Guilford, J. P., & Zimmerman, W. S. *The Guilford-Zimmerman Temperament Survey: Manual of instructions and interpretations*. Beverly Hills, Calif.: Sheridan, 1949.

Guilford, J. P., & Zimmerman, W. S. Fourteen dimensions of temperament. *Psychol. Monogr.*, 1956, **70**(10, Whole No. 417).

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187–200.

Klugh, H. S., & Bendig, A. W. The Guilford-Zimmerman Temperament Survey and intelligence. *Psychol. Newsltr., NYU*, 1959, **10**, 96–97.

Lovell, C. A study of the factor structure of thirteen personality variables. *Educ. psychol. Measmt.*, 1945, **5**, 335–350.

Thurstone, L. L. The dimensions of temperament: Analysis of Guilford's thirteen personality scores. *Psychometrika*, 1951, **16**, 11–20.

Wrigley, C. F., & Neuhaus, J. O. The matching of two sets of factors. *Amer. Psychologist*, 1955, **10**, 419.

(Received April 9, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

SIZE–DISTANCE JUDGMENT IN ORGANIC MENTAL DEFECTIVES [1]

NOËL JENKIN
University of Sydney
AND SALLY A. MORSE
University of Virginia Hospital

Koffka's (1935) discussion of the constancies gave a controversial status to the question concerning progressive development of size constancy in childhood. Of recent years, empirical evidence has accumulated to resolve this controversy. Lambercier (1946a, 1946b), Piaget and Lambercier (1943, 1951, 1956), and Zeigler and Leibowitz (1957) all agree as to the existence of a developmental trend in objective size–distance judgments, which reaches nearconstancy or overconstancy in adulthood.

The present study focuses upon effects of the type of impaired development which is due to organic brain damage. It was expected that organic mental defectives, due to their impaired organizing and integrating abilities, would make judgments which fall short of size constancy, relative to the judgments of normal persons of equivalent age. The defective individuals would thus occupy a place on the developmental curve which is normally marked by a younger age group.

Since an obtained difference in mean size-matches between organic mental defectives and normals could be attributed to the difference in mental age between these groups, it was planned to use a familial mentally retarded group as a control. The familials would have mental ages similar to those of the organics, but would lack the factor of brain damage which the present hypothesis specifies as the cause of judgments atypical for a given chronological age. Success in judging the physical size of a distant object is considered to be an ability developed through normal visual experience and not primarily dependent on intelligence. It was therefore expected that the familial group would make size–distance judgments which are closer to those of the normal group than to those of the organic group.

A supplementary aim of the investigation was to study the trends of the size-matches when the comparison object was presented at three different distances. Previous work by Jenkin (1957, 1959) has shown that with normal adult *Ss*, objective size-matches of a standard at a constant distance become progressively reduced as the comparison object recedes from the point of observation. This tendency, which has been interpreted as an increase of apparent size with increasing distance, was thought to be probably a product of visual learning. If this is so, an interaction of groups by distances would be found when the organics are compared with other groups. The expected direction of the relationship was towards organics showing less difference between distances than either normals or familials.

METHOD

Subjects. Four groups of *Ss* were used, each group numbering 16 cases and including both sexes. Cases of gross visual defect were excluded from each group. The organic group consisted of mentally retarded residents of the Training School at Vineland and the Vineland State School. Each member of this group had been classified in the records of the respective institution as brain damaged from causes occurring prenatally or early in life. The median age of this group was 18 yr. 6 mo., and the median men-

[1] This study was conducted when the authors were at the Training School at Vineland. Thanks are due to the superintendent and staff of the Vineland State School, whose cooperation made possible collection of some of the data.

tal age (Binet, or equivalent) was 7 yr. 10 mo. The familial group was obtained from the same institutions. Its members were classified in the records as mentally retarded cases of familial etiology. The median age was 15 yr. 9 mo., and the median mental age (Binet, or equivalent) was identical with that of the organic group, 7 yr. 10 mo. The normal adolescent group consisted of students from a nearby junior high school who agreed to serve as *Ss* for monetary compensation. The median age was 14 yr. No mental ages were available, but it was ascertained that the students were all achieving at school at a rate normal for their age. The normal adult group consisted of employees of the Training School at Vineland, most of whom were young professional personnel. The median age was 23 yr.

Apparatus. The apparatus consisted of white cardboard squares. The standard card, 4 in. square, was attached by a steel wire to a ringstand and was placed directly in front of *S* at a distance of 320 in. from his eyes. The comparison series was a set of squares ranging from 2½ in. to 6 in. by steps of ¼ in. They were individually displayed at the same height as the standard upon a platform which was situated on *S*'s right. Concealed pins and a slight depression, invisible to *S*, served to maintain the comparison cards in vertical position in the middle of the platform. The angular separation between the standard and the comparison was 87°. Homogeneous fields against which both stimuli were viewed were provided by wooden screens painted a dull green. Little daylight entered the room, and the stimuli were illuminated by 75-w. floodlights placed so as to give an illumination reading of 11 foot candles (measured by a Macbeth Illuminometer) from a correctly positioned square.

Design. Each *S* judged the size of the standard with the comparison object placed at three different distances: 20 in., 80 in., and 160 in. Two trials were conducted at each distance, one an ascending and the other a descending series. Each group of *Ss* (organic, familial, adolescent, adult) was divided into two subgroups, A and B. Subgroup A experienced the different distances in the order 80 in., 160 in., 20 in.; and Subgroup B experienced the order 20 in., 160 in., 80 in. Half the members of each subgroup began with an ascending series and the other half began with a descending series.

Procedure. The height of the stimulus objects and *S*'s eye level were properly matched. The platform was placed at the distance appropriate for viewing the comparison object on the first trial. *S* was then instructed about the nature of the judgment required. He was asked to respond to each comparison card with an answer of "smaller," "the same," or "bigger." It was explained that "the same" meant that the comparison card would exactly match the standard in its physical dimensions. The concepts of "smaller" and "bigger" were illustrated with demonstration cards, so chosen as to lie outside the range of comparison cards used in the experiment. The concept of "equality" was illustrated with a pair of cards 3 in. square, held in *E*'s hands, then superimposed and rotated before *S*'s eyes. Two practice trials were then conducted, one ascending and the other descending. A pilot study had indicated that, with retarded *Ss*, at least this much experience was necessary for the thresholds to attain reasonable stability. Data from the practice trials were therefore excluded from the analysis. *S* was required to look first at the standard and then at the comparison, keeping his head erect and in the same position for each judgment. He was not prevented from repeatedly looking back and forth if he wished to do so. Within predetermined limits, *E* varied the size of the card with which each series began, and also randomly switched from one set of comparison cards to a second identical set.

Results

Table 1 shows the mean size-matches at each of the three distances for the four groups of *Ss*. It will be seen that the grand mean for the organic group is less than 4 in., which latter value would represent perfect constancy. The grand mean for every nonorganic group is higher than 4 in.

The results of an analysis of variance of the data are shown in Table 2. Variability among the four groups is statistically significant ($P < .01$). The three distances also represent a significant source of variance. The specific hypotheses of the experiment are answered by t tests of the significance of the difference between pairs of groups. With 30 df, the value obtained for the organics vs. familials comparison was $t = 2.32$, $p < .05$. A test of the organics vs. adolescents comparison yielded $t = 2.63$, $p < .02$, and between the organics and adults, t reached a value of 4.06, $p < .001$. All comparisons of nonorganic pairs of groups, as expected,

TABLE 1

Mean Size-Matches, in Inches, of a 4-in. Standard[a] by 4 Groups of Subjects

Group	Distance of Comparison Object			Grand Mean
	20 In.	80 In.	160 In.	
Organic defectives	3.777	3.805	3.746	3.776
Familial defectives	4.178	4.250	3.982	4.137
Normal adolescents	4.246	4.195	4.113	4.185
Normal adults	4.605	4.355	4.262	4.407

[a] The standard was 320 in. from the point of observation.

TABLE 2

ANALYSIS OF VARIANCE OF SIZE-MATCHES

Source	*df*	*MS*	*F*
Groups (G)	3	210.25	5.66*
Order (O)	1	103.55	2.79
G × O	3	49.15	1.32
Ss	56	37.16	
Distances (D)	2	33.57	7.81**
D × G	6	7.96	1.85
D × O	2	5.51	1.28
D × G × O	6	3.57	0.83
Residual	112	4.30	

* $P < .01$.
** $P < .001$.

showed nonsignificant differences, p being greater than .05 in every case.

It is noteworthy that the organic group made a mean size-match significantly smaller than that of the familial group, while the size-match of the latter group did not differ significantly from those of either of the normal groups. The atypical performance of the organics cannot be attributed to low measured intelligence, since the two retarded groups were equated on this variable.[2]

The effort to match the ages of the organics, familials, and normal adolescents was not entirely successful. The unavoidable error of having an organic group with too high a median age is, however, in a direction which weakens the chances of a positive finding. Both the familial and the normal adolescent groups are, on the average, younger than the organic group and, on the developmental hypothesis, this would tend to minimize any differences which may be due to organicity. Nevertheless, our hypothesis concerning the relationship of the organics to the other groups is strongly supported by the data.

An examination of Table 1 shows that the trend of mean size-matches among the three different distances of the comparison appears as expected for the normal groups, i.e., towards smaller matches as distance increased. Also as expected, there is less variability between distances in the case of the organic group. Table 2 indicates, however, that the differences among these trends are not great enough to produce a significant groups × distances interaction.

[2] A further test of the possible role of intelligence was made by computing the rank correlations between mental ages and mean size-matches. Nonsignificant values of −.22 for the organic group and −.05 for the familial group were obtained.

DISCUSSION

The present study has demonstrated that organic mental defectives tend to underestimate the size of a distant object. Normal persons tending to be younger than the defectives are inclined to overestimate the size of a distant object, and so are normals who tend to be older than the defectives. The difference between the performances of organics and normals cannot be attributed to the difference in Binet mental age, as shown by the performance of a control group of familial etiology.

These results, which support the main hypothesis of the investigation, may be usefully compared with data from a study of normal children and adults by Piaget and Lambercier (1951). A group of children, aged from 7–8 yr., made a mean objective match which fell short of size constancy. The next group, aged from 8–10 yr., slightly exceeded size constancy, and the three remaining groups respectively gave average values which showed a progressive overestimation as age increased. The direction and extent of the difference between Piaget and Lambercier's adolescent (12–14 yr.) and adult groups resemble what was found from the comparable groups of the present study. Our organic group, however, behaved like the young children (7–8 yr.) of Piaget and Lambercier, even though our cases were very much older.

It appears reasonable, therefore, to interpret the behavior of our organic group as indicating immaturity of size–distance judgment, due to brain damage occurring prenatally or early in life. A firm conclusion of this type is not at present possible, since the present authors did not have the opportunity to study a group of individuals who had suffered recent brain damage. A future experiment might determine whether or not such persons behave like normals or like young children in the size–distance judging situation, and the role of learning would thus be further clarified.

In the meantime, the authors consider that the present results are generally consistent with other work on perception in organic mental defectives, summarized by Strauss and Kephart (1955) and by Gallagher (1957), and might usefully be interpreted along broadly similar lines. This is to the effect that prenatal or early brain damage impairs the development of integrative cerebral mechanisms which are necessary for normal perceptual functioning (Strauss & Kephart, 1955, pp. 57–58). Organics thus tend to show features such as a fragmentary response to a complex pattern of stimulation, inadequate perceptual discrimination, and a failure of synthesis in a situation where several sources of information are normally combined.

It is the latter feature which in particular seems to have been demonstrated in the present study. It cannot be pretended that a fully satisfactory theory of size constancy exists, but it is known that its manifestation depends upon the apprehension of various depth cues and the integration of these with other sources of information. The proximal stimulus size and also identification and categorization of the object displayed are probably included among these sources. The chronically impaired ability to integrate several sources of information and the consequent weakness of the information which is given by experience are factors probably responsible for the tendency of the organic defective to rely more heavily than other groups upon the proximal stimulus size.

It is of interest to note in this connection that "normal" perception is not necessarily the most veridical perception. Table 1 indicates that the organic group was, on the average, more successful than the adult group in matching the objective size, and the same statement may be made about the young children in Piaget and Lambercier's study. As pointed out by Jenkin and West (1958), in connection with the size–weight illusion, the gross degree of error measured in some perceptual situations is not an index of perceptual inefficiency; rather, it is the opposite. This point of view leads us to interpret the behavior of the organics as tending to be "stimulus bound," whereas the normals, by virtue of their superior organizing ability and more adequate fund of experience, tended more to respond to the total situation, including its illusory [3] aspects.

Summary

Binocular objective size-matches of a standard 4-in. square situated 320 in. from the observer were obtained at three different distances of the comparison object. Under conditions of strong depth cues and ample illumination, judgments were made by four groups of *Ss* ($N = 16$ per group): (*a*) familial type mental defectives, (*b*) brain damaged mental defectives, (*c*) normal adolescents, (*d*) normal adults. As hypothesized, the organics made matches significantly closer to the visual angle than those of the other groups. There was no significant difference between the familials and the normals. As expected, variation among the different distances was smaller in the case of the organic group, but this trend was not statistically significant. The data were discussed in terms of perceptual learning, and it was suggested that the performance of the brain damaged cases is due to chronically impaired ability to integrate several sources of information.

References

Gallagher, J. J. *A comparison of brain injured and non-brain-injured mentally retarded children on several psychological variables.* Lafayette, Ind.: Child Developm. Publ., 1957.

Jenkin, N. Effects of varied distance on short-range size judgments. *J. exp. Psychol.*, 1957, **54**, 327–331.

Jenkin, N. A relationship between increments of distance and estimates of objective size. *Amer. J. Psychol.*, 1959, **72**, 345–363.

Jenkin, N., & West, Noël. Perception in organic mental defectives: An exploratory study. I. The size-weight illusion. *Train. Sch. Bull.*, 1958, **55**, 5–10.

Koffka, K. *Principles of Gestalt psychology.* New York: Harcourt, Brace, 1935.

Lambercier, M. La constance des grandeurs en comparaisons sériales. *Arch. Psychol., Genève,* 1946, **31**, 79–282. (a)

Lambercier, M. La configuration en profondeur dans la constance des grandeurs. *Arch. Psychol., Genève,* 1946, **31**, 287–324. (b)

[3] All *Ss* were shown the standard at close range (20 in.) after completion of the experiment. A large majority of the normal adolescents and adults remarked that it looked smaller than was thought under conditions of the test.

PIAGET, J., & LAMBERCIER, M. La comparaison visuelle des hauteurs à distances variables dans le plan frontoparallèle. *Arch. Psychol., Genève,* 1943, **29,** 175–253.

PIAGET, J., & LAMBERCIER, M. La comparaison des grandeurs projectives chez l'enfant et chez l'adulte. *Arch. Psychol., Genève,* 1951, **33,** 81–130.

PIAGET, J., & LAMBERCIER, M. Grandeurs projectives et grandeurs réeles avec étalon éloigné. *Arch. Psychol., Genève,* 1956, **35,** 257–280.

STRAUSS, A. A., & KEPHART, N. C. *Psychopathology and education of the brain-injured child.* Vol. II. *Progress in theory and clinic.* New York: Grune & Stratton, 1955.

ZEIGLER, H. P., & LEIBOWITZ, H. Apparent visual size as a function of distance for children and adults. *Amer. J. Psychol.,* 1957, **70,** 106–109.

(Received April 13, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

IMPULSE REPRESSION AND EMOTIONAL ADJUSTMENT

HARRY GRATER [1]

Michigan State University Counseling Center

Theories of personality development have emphasized what behavior the individual considers acceptable or right and what behavior he considers unacceptable or wrong. These judgments of rightness or wrongness form important parts of psychological concepts such as value systems, super ego, ego-ideal and self concept. The basic premise is that while very young the child learns that certain types of behavior resulted in parental approval, while other forms of behavior bring disapproval. Since the young child's security and sense of well being are so closely bound to parental evaluation, he soon learns to avoid the disapproved forms of behavior and eventually to repress impulses to such acts. The child experiences anxiety if he behaves in a nonacceptable fashion or even if impulses to such behavior are experienced.

Since the above position places such strong emphasis on conflicts between impulses to behavior and forces acting to inhibit behavior, it could be called an impulse repression theory of anxiety.

The experimental evidence concerning the validity of the impulse repression position is contradictory. Stone and Landis (1953) found that high school seniors who were raised in a democratic family structure experienced fewer conflicts with their families than did those raised in an autocratic family. Watson (1934) studied the effects of permissive vs. strict home discipline and found that subjects raised under strict home discipline reported more indications of emotional conflict; however, Watson (1957) in a later study found no difference in the emotional adjustment of children raised in strict and permissive families. Brown (1942) found no relationship between the attitudes of parents toward the child's behavior and the child's adjustment. Lasko (1952) reported that children raised by laissez faire parents gave indications of emotional maladjustment.

Two factors limit the effectiveness with which one can use the above empirical data to evaluate the impulse repression theory of anxiety. First, as previously indicated, the above studies are contradictory. Second, they have tended to emphasize the democratic-autocratic continuum. It is quite possible for a family to have an autocratic structure and yet allow the child relatively free expression of certain basic impulses. Thus, determining that an autocratic family structure results in greater maladjustment does not indicate that this maladjustment resulted from the repression of basic impulses.

Since the impulse repression theory already presented depicts anxiety as resulting from conflicts between impulses and repression, the child who has been taught a large number of behavior restrictions by the significant people in his life should experience greater anxiety than the child brought up under more liberal conditions. The above should be true for there would be more opportunity for the child with a high set of behavior standards to experience conflicts between impulses and standards, as well as between behavior and standards. If the child with the high set of behavior standards experiences more conflict, he should have also developed more symptoms of these conflicts, that is, more neurotic symptoms.

This research was designed to explore the following questions which are related to the impulse repression theory of anxiety already presented. Can subjects who report a high number of neurotic symptoms be distinguished

[1] This paper is based upon a doctoral dissertation completed at Pennsylvania State University. The writer should like to acknowledge the advice and suggestions of William U. Snyder under whose direction this study was done.

from subjects who report a smaller number of neurotic symptoms on the following factors:

(*a*) The number of moral and ethical standards of behavior held by the subject?
(*b*) The frequency with which they engaged in behavior involving a moral or ethical judgment?
(*c*) The number of moral and ethical behavior standards held by one of the significant adults in the subject's life as perceived by the subject?
(*d*) The number of moral and ethical behavior standards held by the significant adult in the subject's life?
(*e*) The frequency with which the subject engaged in behavior which he considered wrong, perceived a significant adult in his life as considering wrong, or is considered wrong by a significant adult in his life.

Methods and Procedures

Single, white, Protestant university females were selected as the population to be used in this research. Mothers were selected as the significant adult in the subject's life.

The Minnesota Multiphasic Personality Inventory was chosen as the instrument to measure emotional adjustment. Since the experimental questions referred to neurotic symptoms, these three scales forming the neurotic triad, hypochondriasis (*Hs*), depression (*D*) and hysteria (*Hy*), were selected as measures of neurotic symptomatology.

In order to obtain items for a scale of behavior standards, a questionnaire was sent to 175 college females. They were asked to list the significant moral or ethical decisions that females of their age were called upon to make. A list of 67 items was compiled from the returns to this questionnaire.

These items were used to form three scales. Each item was contained on all scales, but the subjects were asked to respond in three different ways. On the first scale, they indicated whether they believed the behavior described would be morally or ethically right or wrong for all females of their age. On the second scale, they were asked to indicate whether they believed their mother would consider their behavior right or wrong for females of their age, and on the third scale, they were asked to indicate whether or not they had engaged in the behavior described. There were called the Daughter Wrong, Perceived Mother Wrong, and Not Done scales, respectively.

These three scales were administered to a sample of 100 college females. The scales were scored by counting the number of items checked wrong on the Daughter Wrong and Perceived Mother Wrong scales, and by counting the items marked as not done on the Not Done scales. The results were used in an item analysis. Responses of subjects falling in the top and bottom quarters were compared for each of the three scales. Phi coefficients were calculated graphically for each item. Those items which discriminated between the top and bottom quarters on one or more scales at the 5% level of confidence were retained. A total of 62 items were included on the basis of this criterion.

The items included on the final scale measured behavior or attitudes involving hostility, heterosexuality, religion, college regulations, and some items less easily described which involve some of the social behavior expected of university females. The following are examples of each of the above types of items:

(*a*) being critical of your parents
(*b*) having sexual relations prior to marriage if engaged
(*c*) changing religious beliefs because of what is taught in a course
(*d*) breaking living unit regulations regarding hours
(*e*) social drinking

Split half reliabilities were calculated for each scale. These corrected reliabilities were as follows: Daughter Wrong .87; Perceived Mother Wrong .84; Not Done, .88.

The MMPI and the Moral Ethical Values Scales were administered to a sample of 136 college females. The sequence in which the Moral Ethical Value Scales were administered was alternated in order to reduce the influence of any particular sequence. All of the subjects included in the population were single, white, Protestant, of average college age who, except for the time spent at college, were living with their mothers at the time of testing. All were volunteers from second or third year courses in psychology or education. They were not asked to put their names on the test. Instead they signed their names to a sign-up sheet and code numbers were used to match tests and subjects. The volunteers did not know the matchings could be made.

The items on the Moral Ethical Values scales were sent to the mothers of these 136 subjects. They were asked to indicate whether they felt that each item described behavior which would be morally or ethically right or wrong for a female of college age. 102 mothers responded and 95 of the responses were usable. The others were discarded for failure to follow directions accurately. The mothers were not asked to sign their names to these scales and the mothers and daughters scales were matched by using code numbers.

The daughters of the 95 mothers whose responses were used in the study were compared with the daughters of the mothers who were not included. The scores received on the neurotic triads of the MMPI and on the three Moral Ethical Value scales were used for these comparisons. Means and critical

ratios were calculated to determine if there were significant differences between the means of these two groups. No significant differences were found between these two groups on the scores they received on either the neurotic triad of the MMPI or on the Three Moral Ethical Value scales.

The subjects received the following scores on the Moral Ethical Values scales:

(*a*) Daughter Wrong (DW). The number of items on this scale which the daughter considered morally or ethically wrong for a female of her age.

(*b*) Perceived Mother Wrong (PMW). The number of items the daughter marked indicating that she believed her mother would consider the behavior being described as morally or ethically wrong for a college female.

(*c*) Not Done (ND). The number of items on the scale which described behavior in which the subject had not engaged.

(*d*) Mother Wrong (MW). The number of items on the scale which the mother indicated she would consider morally or ethically wrong for a female of college age.

(*e*) Daughter Behavior Conflict (DBC). The frequency with which the daughter had engaged in behavior which she indicated was morally or ethically wrong.

(*f*) Perceived Mother Behavior (PMBC). The frequency with which the daughter had engaged in behavior which she perceived her mother as considering morally or ethically wrong.

(*g*) Mother Behavior Conflict (MBC). The frequency with which the daughter had engaged in behavior which the mother considered morally or ethically wrong.

The subjects were divided into groups on the basis of the scores they received on the MMPI. All of the subjects rated in the top third of the sample on the *Hs* scale were placed in one group called the *Hs* group. Subjects who fell into the top third on the *D* scale were placed in a *D* group, and subjects who rated in the top third on the *Hy* scale were placed in a *Hy* group. It was possible for subjects to be included in more than one neurotic group if they had received high ratings on more than one scale. The remaining subjects, none of whom had been included in a neurotic group, were formed into a nonneurotic group.

Table 1 shows the mean *T* scores and *T* score ranges received by the neurotic and nonneurotic groups on the MMPI.

Divisions into top 3rd and bottom 2/3rds could never be exact because subjects received scores which were at the cut-off point between the upper third and bottom two thirds. This was resolved by including all subjects who tied at this cut-off point in the nonneurotic group. The number of subjects in each group is as follows: *Hs*, 35; *D*, 39; *Hy*, 44; Nonneurotic, 67. Only 95 of the mothers of the total of 136 subjects were used in the study. Thus, the size

TABLE 1

MEAN *T* SCORES (ROUNDED) AND *T* SCORE RANGES RECEIVED BY THE NONNEUROTIC AND NEUROTIC GROUPS ON THE NEUROTIC TRIAD OF THE MMPI

Group	Mean *Hs* *T* Score	*Hs T* Score Range
Nonneurotic	46	28–52
Hypochondriacal	59	53–72
Nonneurotic	43	32–51
Depressive	60	52–76
Nonneurotic	50	40–57
Hysterical	65	58–80

of each of the above groups was reduced when the MW scores were used. The number in each group was reduced to: *Hs*, 24; *D*, 24; *Hy*, 31; nonneurotic, 48.

Means for these four groups were calculated for the scores received on the Moral Ethical Values scales. Each mean score received by the neurotic groups was compared with the means received on the same scale by the nonneurotic group. Critical ratios were calculated to determine whether the differences were significant.

RESULTS

Table 2 shows the mean scores received by the neurotic and nonneurotic on the Moral Ethical Values scales. Critical ratios have been calculated for the difference between

TABLE 2

MEANS AND THE CRITICAL RATIOS OF THE DIFFERENCES BETWEEN THESE MEANS OF THE SCORES RECEIVED BY THE NONNEUROTIC AND NEUROTIC GROUPS ON THE MORAL-ETHICAL VALUES SCALES

Group	DW	ND	PMW	MW
Nonneurotic	37.04	28.60	47.81	51.48
Hypochondriacal	36.01	28.58	44.72	47.81
Critical ratio	.53	.00	1.82	1.94
Nonneurotic	37.04	28.60	47.81	51.48
Depressive	35.31	28.15	45.43	47.87
Critical ratio	.97	.21	1.23	1.82
Nonneurotic	37.04	28.60	47.81	51.48
Hysterical	34.75	28.66	44.16	47.00
Critical ratio	1.29	.00	2.50*	2.73**

* Significant at 5% level of confidence.
** Significant at 1% level of confidence.

these means in order to evaluate the significance of the differences.

Table 2 shows that there are no significant differences between the DW scores of neurotic and nonneurotic groups. A critical ratio of only 1.29, which is not large enough to be significant, was calculated for the difference between these two means. Behavior standards then did not distinguish the neurotic from the nonneurotic group. The slight differences which were found were in the direction of the neurotic groups having lower scores on the DW scale.

There were no significant differences between the ND scores of the neurotic and nonneurotic groups. Thus, there were no differences in the reported behavior of these two populations in situations involving a moral or ethical judgment.

The subjects in the *Hy* group received lower PMW scores than did the subjects in the nonneurotic groups. This difference reached the 5% level of significance. The *Hs* and *D* groups also perceived their mothers as considering fewer items wrong, but these differences were not large enough to be significant; although in the case of the *Hs* group, the difference did reach the 10% level of confidence.

In all three comparisons, the mothers of subjects in the neurotic groups considered fewer items wrong than did mothers of the nonneurotics. All of these differences reached the 10% level of significance, and when the mothers of the subjects in the *Hy* group were compared with mothers of the subjects in the nonneurotic groups, this difference reached the 1% level of significance.

Thus, both the mothers and the perceived mothers of the hysterical group considered a significantly fewer number of the items wrong than did the mothers of the nonneurotics. The same tendency was found when PMW and MW scores received by the other neurotic groups were compared with the nonneurotic group, but the differences were not large enough to be significant above the 10% level of confidence.

The data in Table 3 are presented in terms of proportions instead of in actual frequencies. This was done because all of the conflict scores are in part dependent upon the DW, PMW, and MW scores. For example, a subject having a high MW score would have more opportunity to come in conflict with the maternal standards than a subject with a lower score. Since there were differences in the DW, PMW, and MW scores of the neurotic and nonneurotic groups, differences between the proportions of conflict rather than an actual count of the number of conflicts give more meaningful results.

TABLE 3

PROPORTION OF THE ITEMS ON THE DW, PMW, AND MW SCALES MARKED AS MORALLY OR ETHICALLY WRONG BUT WHICH DESCRIBED BEHAVIOR IN WHICH THE SUBJECT HAD ENGAGED

Group	DBC	PMBC	MBC
Nonneurotic	35	46	49
Hypochondriacal	36	46	47
Critical ratio	.10	.28	.15
Nonneurotic	35	46	49
Depressive	35	48	51
Critical ratio	.00	.20	.16
Nonneurotic	35	46	49
Hysterical	35	43	45
Critical ratio	.00	.30	.35

Note.—Critical ratios have been calculated to determine if these conflict scores are significantly different for the neurotic and nonneurotic groups.

Table 3 shows no significant differences between the conflict scores received by the neurotic and nonneurotic groups. This indicates that the neurotic subjects do not tend to adhere more closely to their own standards, their perceived maternal standards, or the maternal standards than do the nonneurotic subjects.

DISCUSSION

These results either do not uphold, or are directly antithical to the impulse repression theory of anxiety which was previously presented. In no instance did a neurotic group receive a higher PW, PMW, ND, or MW score indicating greater inhibition of behavior than did the nonneurotic group. When differences that approached or reached a significant level were found, the neurotic groups received lower scores indicating less behavior restriction. Nor did the conflict scores indicate that the neurotics were more controlled

by their own standards, the perceived mother's standards, or the mother's standards than were the nonneurotics.

The basic question raised by this research is "Why were low PMW and DW scores associated with Neurosis?" Perhaps an explanation can be found in Fenichel's (1945) discussion of what he refers to as the heteromerous super ego or in Mowrer's (1950) discussion of super ego repression. Both Fenichel and Mowrer have emphasized the effects of the individual's confusion about standards of behavior. It is possible that the liberal PMW and MW scores received by the mothers and the perceived mothers were actually indications of confusion rather than of liberality, and further that confusion is related to maladjustment.

In order to evaluate this confusion hypothesis a confusion score was calculated and correlated with the *Hy* score of the MMPI. The *Hy* scale was selected because it was most closely associated with low PMW and MW scores. The confusion score was calculated by counting the number of times the daughter had misperceived the mother's standards. The Pearson Product-Moment Correlation between the frequency with which the daughter had misperceived her mother's standards and hysteria was .30. This correlation is significant at the 1% level of confidence. Thus the daughter's confusion about maternal standards was correlated with a measure of neuroticism.

This research introduces questions which need further exploration. The first step might be a replication of the experiment reported here. If at all possible, neurotic groups should be selected which are more deviant than the subjects included in this study. Also, subjects should be allowed to respond to the items with a scaled response rather than with a dichotomous response. Such scaling would allow an evaluation of the confusion hypothesis which was already presented.

Summary

The objective of this research was to test the impulse repression theory of anxiety. A scale of behavior standards called the Moral Ethical Value scale and the MMPI were administered to a population of 136 single, white, Protestant college females. On the first scale of the Moral Ethical Values scale, the subjects were asked to indicate whether they felt the behavior described by each item would be morally or ethically right or wrong for a college female. On the second scale they were asked to indicate whether they felt their mothers would consider the behavior described as morally or ethically right or wrong for a college female, and on the third scale they were asked to indicate whether they had engaged in the behavior being described. Scales were also sent to the mothers who were asked to indicate whether they felt the behavior described would be morally or ethically right or wrong for a college female.

Subjects were divided into depressive, hysterical, hypochrondriacal and nonneurotic groups on the basis of scores received on the neurotic triad of the MMPI.

The neurotic and nonneurotic groups were compared on the scores they received on the Moral Ethical Values scales. These scales gave a measure of the mothers' and daughters' evaluation of behavior in situations involving moral or ethical judgment. The results showed no significant differences between the neurotic and nonneurotic group in the number of items on the scale that they indicated were morally or ethically wrong, nor was the behavior of the neurotic group different from the behavior of the nonneurotic group in situations involving a moral or ethical judgment. The neurotic daughters tended to perceive their mothers as considering fewer items wrong than did the nonneurotics. Also the mothers of the neurotics tended to consider fewer items wrong than did the mothers of the nonneurotics. Both the perceived mothers and the mothers of the *Hs* subjects considered significantly fewer items wrong than did the mothers of the nonneurotic subjects. There was no difference in the proportion of time the neurotics and nonneurotics engaged in behavior which conflicted with their own, the perceived mothers, or the mother's behavior standards.

The results of this research are antithical to the impulse repression theory of anxiety. From this theory one would predict greater restriction of behavior to characterize the

neurotic groups. In some comparison the direct opposite of these predicted results was found.

The hypothesis was presented that liberality was actually a reflection of confusion about moral and ethical standards. In order to evaluate this hypothesis, the frequency with which the daughter misperceived her mother's standards was correlated with the hysteria scale of the MMPI. A significant correlation of .30 was calculated. This tended to support a confusion hypothesis.

REFERENCES

BROWN, F. An experimental study of parental attitudes and their effect upon child adjustment. *Amer. J. Orthopsychiat.*, 1942, **12**, 224–230.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FREUD, S. *The problem of anxiety.* New York: Norton, 1936.

GUILFORD, J. J. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1942.

LASKO, JOAN. Parent-child relationships. *Amer. J. Orthopsychiat.*, 1952, **22**, 300–306.

MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.

STONE, CAROL, & LANDIS, PAUL H. An approach to authority patterns in parent-teen age relationships. *Rur. Sociol.*, 1953, **18**, 223–242.

SULLIVAN, H. S. *Conceptions of modern psychiatry.* Washington, D. C.: William Alanson White Psychiatric Foundation, 1947.

WATSON, G. A comparison of the effects of lax versus strict home training. *J. Soc. Psychol.*, 1934, **5**, 102–105.

WATSON, G. Some personality differences in children related to strict or permissive parental discipline. *J. Psychol.*, 1957, **44**, 227–249.

(Received July 31, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

CRITICAL FLICKER FREQUENCY AND THE SPIRAL AFTEREFFECT WITH PROCESS AND REACTIVE SCHIZOPHRENICS[1]

JOSEPH M. McDONOUGH[2]
Michigan State University

Considerable theory and some research suggest that at least certain schizophrenics suffer from a disease that is primarily organic in nature, perhaps of a central nervous system locus. Traditionally it has been the tacit assumption of many workers in the area that schizophrenia has an organic etiology, most likely involving a disturbance of the brain itself (Bleuler, 1950; Noyes, 1953). The number of published findings to support such an hypothesis are far outnumbered by articles which merely speculate on such a possibility without any attempt at experimentation or controlled observation. A few studies nevertheless do offer some support for the assumption that brain damage is *the* important factor in schizophrenia. Evidence from biopsies obtained from 70 cases of dementia praecox led Bateman and Papez (1951) to suggest that changes in nerve cells of the prefrontal cortex are due to pathogenic organisms. From an examination of 1087 records, Kallmann (1938) concluded that "the organic process of schizophrenia in the brain . . . attacks the cortex almost exclusively and usually leaves the internal parts of the brain intact." Kennard and Levy (1952) found abnormal EEGs, often indicative of brain lesions, significantly more often in the more severe cases of schizophrenia, especially those resembling dementia praecox. Winkelman and Book (1948) concluded on the basis of 10 autopsies of schizophrenics that gross brain appearance did not differ from normals but that microscopic evidence did suggest the diagnosis.

Widespread dissatisfaction with the traditional Kraepelinian categorization of schizophrenia or dementia praecox has led to increased effort in recent years to refine the concept. One suggestion has been to view the disorder as including two polar types. At one extreme is the patient who has an early and insidious onset of psychosis with absence of precipitating stress, an inadequate prepsychotic personality with a tendency to withdraw from all human contacts, clinical symptoms showing a flat or dull affect, and relative absence of confusion. This syndrome has been called *process* schizophrenia. *Reactive* schizophrenia on the other hand is the term given to describe the case where there is typically an abrupt and stormy onset of the psychosis usually attributable to a logical and significant stress situation. Here the premorbid personality may be described as normal or neurotic, rather than schizoid, with perhaps some degree of outgoingness present and a clinical picture likely to include severe confusion and many affective components. Attempts to classify schizophrenic samples into the two categories or to view the individual patient as occupying a position on a process-reactive continuum have increased in number and a certain degree of success has been achieved in differentiating the groups experimentally (Becker, 1956; Kantor, Wallner, & Winder, 1953; King, 1958). One investigation by Brackbill and Fine (1956) sought to

[1] This research is based on a doctoral dissertation done at Michigan State University. The author wishes to express his appreciation to M. Ray Denny, Gerald F. King, S. Howard Bartley, Alfred Dietze, and G. M. Gilbert for their many contributions to this thesis. Particular thanks are due also to Stewart G. Armitage and his staff at the VA Hospital, Battle Creek, Michigan for their continuous cooperation in making this study possible.

[2] Now at the Veterans Administration Hospital, Coral Gables, Florida.

relate the concepts specifically to central nervous system pathology. Although it was impossible to differentiate the process from the organic group on the basis of Rorschach indices of brain damage, it was observed that both of these groups had significantly more individuals with five or more of the "organic signs" than did the reactive group. The authors suggested, therefore, that "the difficulty in differential diagnosis of some kinds of schizophrenics and organics results from the involvement of central nervous system pathology in process schizophrenia."

Two perceptual techniques of "brass instrument" psychology have been found fairly effective in differentiating brain damaged from nonorganic groups: critical flicker frequency (CFF) and the Archimedes Spiral. Numerous studies have substantiated the fact that brain damaged patients have significantly lower CFF thresholds than do nonorganic patients (Ross & Reitan, 1955; Teuber & Bender, 1948; Werner & Thuma, 1942). In psychiatric patients with functional disorders CFF has not been investigated as thoroughly, and most studies are open to methodological criticisms including the small number of patients involved. In two of the better controlled studies, however (Irvine, 1954; Ricciuti, 1948), the preponderance of evidence indicated that schizophrenic patients have significantly higher CFF thresholds than do brain damaged patients and that they do not differ significantly from normal individuals.

Visual fixation by *S* upon a slowly rotating Archimedes Spiral produces the illusion of expansion or contraction of the disc depending on the direction of rotation. When the spiral is brought to a sudden stop the perceiver experiences the illusion just the reverse of that during rotation, i.e., a negative aftereffect. Although the extremely high discrimination achieved by Price and Deabler (1955) in differentiating brain damaged from nonorganic patients has not been replicated, subsequent studies generally have found that significant differences do exist between the two groups and that the technique is effective to some extent as a clinical test of organicity (Gallese, 1956; Page, Rakita, Kaplan, & Smith, 1957).

On the assumption, therefore, that process schizophrenia involves central nervous system pathology and that this is specifically cortical in nature, it was hypothesized that process schizophrenics and brain damaged patients would have significantly lower CFF thresholds and would fail to perceive the spiral aftereffect significantly more often than would reactive schizophrenics and normals.

Method

Apparatus

The CFF was tested by means of an episcotister viewed monocularly by *S*. Visual angle of the target was 2 degrees 28 minutes with an intensity of .085 candles per square foot. A shutter which could be lowered over the target aperture allowed *E* to terminate the stimulus immediately at the end of each flicker trial. Power for the episcotister was supplied by a Cenco 115-volt, 60-cycle motor. Rate of rotation for the disc was controlled manually by *E*'s turning a small knob on the motor. Four equal sized sectors were cut in the disc in a manner to produce a light-dark ratio of 1.0. An automobile speedometer was used to measure the speed of the spinning disc and calibration by a Strobotac proved linear.

The Archimedes Spiral was made from a cardboard disc, off-white in color and 8 inches in diameter. A black stripe, one-half inch wide, was painted on the cardboard so that it made one and three quarter turns from the center until it tapered completely off the outside edge of the disc. Speed of rotation was kept constant at 100 rpm.[3] When rotated, an illusion of the disc contracting or drawing into the center was created in the typical viewer. On stopping the disc suddenly the "normal" experience was to perceive the spiral as expanding.

Subjects

Eighty hospitalized male veterans between 19 and 45 years of age with good visual acuity were placed in four groups of 20 *S*s each. The organic group consisted of individuals with *known* brain damage, chosen on the basis of their diagnosis as determined by neurological findings of the medical staff, case history material, and presenting symptomatology. Diagnoses included multiple sclerosis, brain trauma, brain tumor, Huntington's Chorea, Parkinson's Disease, CVA, and intercranial infection. Mean age was 35.2 years with patients averaging 9.3 years of schooling. The normal group was made up of 20 veterans from a general medical and surgical hospital and included every man on a surgical ward under the age of 45 with no history of emotional difficulties necessitating hospitalization or outpatient treatment. Diagnoses included postoperative fractures, appendectomies, torn cartilage, hemorrhoids, and sacroiliac

[3] Some unpublished evidence appears to indicate that size of disc and speed of rotation affect *S*'s perception of the afterimage.

strain. This group averaged 31.1 years of age with a mean educational level of 10.7 years.

The third and fourth groups were made up of diagnosed schizophrenics from an NP hospital. One hundred and sixty-one schizophrenic case histories were examined. From these a total of 76 were selected on the basis of their having: a case history with sufficient material to allow rating; no evidence either behaviorally or from the record of any head injury or long period of unconsciousness; a background free from repeated NP hospitalization during the previous 10 years; and definite evidence from the history of gross schizophrenic symptomatology, e.g., hallucinations, delusions, ideas of reference, and bizarre behavior. These 76 patients were then rated on Becker's (1955) revision of the Elgin Prognostic Scale by two clinical psychologists. Interrater reliability was indicated by a Pearson r of .89. The two numerical ratings on each patient were then averaged and the resulting 76 scores ranked on a continuum. The 20 patients receiving the lowest point totals were selected as being most "reactive," and those 20 with the highest scores most "process." The decision to use Becker's rating scale was based on the belief that it offered a greater degree of exactness in the ratings, permitted a grading along the entire length of the continuum rather than a forced choice dichotomy or the creation of a middle or mixed group, and appeared to fit more easily into a theoretical framework such as Bellak's (1949) where schizophrenics are seen as ranging from almost completely organic to almost completely psychogenic.

The *process* group had a mean age of 26.6 and an educational level of 10.7 years, while the *reactive* group averaged 33.6 years of age and had 10.8 years of schooling. Mann-Whitney U tests revealed no significant differences among the four groups in terms of educational level, but significant differences in age were obtained. Process schizophrenics were significantly younger than the other three groups ($p < .001$), a characteristic consistent with the early insidious onset associated with this group. Organics were significantly older than normals ($p < .05$), which may be partly attributable to the progressive nature of some of the organic disorders, allowing the patients to remain unhospitalized until somewhat later in life.

Procedure

All *Ss* were informed that they were to undergo two "special eye tests." With the Archimedes Spiral, *S* were seated 8 ft. from the face of the disc with the room under ordinary illumination. Instructions used were those of Gallese (1956) in which the patient was questioned as to his perception of the rotating disc. Four 30-sec. trials were given at the end of which the disc was stopped abruptly. *S* was then asked to describe what he "saw." A scoring system devised by Price and Deabler (1955) was used to record *S*'s response. If his answers suggested that he perceived the aftereffect, he was given a score of 1 for that trial. If he did not report the illusion of expansion, a score of 0 was recorded. *S*'s final score was based on all four trials. A point total of 4, therefore, indicated that *S* perceived the aftereffect on all four trials, whereas a score of 0 meant that he failed to perceive the effect on any of his trials.

In the case of CFF, *S* was carefully instructed on how to view the patch and given the following directions:

> The spot you see there [i.e., the test patch] is going to flicker off and on. Stare at it steadily and don't blink. After a while the light will no longer appear to be flickering. Instead, as you stare at it, it will gradually appear to become a steady light. As soon as it looks as though it is no longer flickering, you say, "Now," . . . and I will turn it off.

The overhead light was then extinguished and the only illumination came from the well-shielded 7½-watt bulb in the rear of the apparatus which permitted dial readings. After one minute of dark adaptation, the shutter was raised allowing *S* to view the test patch. *E* manually increased the disc speed at a constant rate (approx. 1 mph per second) until *S* reported the appearance of a steady light. A total of 10 successive ascending trials were taken with about 30 seconds between trials. Possible auditory cues from the episcotister were masked by an electric fan whose blades struck a small piece of cardboard attached to the fan guard. After considering the first three trials as practice, mean flicker threshold was determined for each of the 80 *Ss*. Patients were found to be quite reliable in their reporting of fusion with only 12 *Ss* obtaining an intertrial *SD* exceeding three quarters of a cps and only one *S* in the entire sample with an *SD* more than 1 cps.

Results

Critical Flicker Frequency

Mean CFF scores for each of the four groups are summarized in Table 1. Since variances were not homogeneous, the Kruskal-Wallis H Test was used. An H value significant at the .001 level of confidence was obtained indicating that the four groups differed in mean flicker threshold beyond chance expectation. Mann-Whitney U Tests run be-

TABLE 1

Critical Flicker Frequencies in Cycles per Second

	Mean	Variance
Organic	13.20	2.30
Normal	16.38	.26
Process	16.43	.73
Reactive	16.67	.94

tween all combinations of the groups in pairs showed that the organic group differed very significantly ($p < .0001$) from each of the other three groups. A significant difference in CFF at the .05 level was found to exist also between normals and reactives. Although mean CFF for the process group was slightly lower than that of the reactive group, the difference was not significant. In order to determine whether a more rigorous dichotomization would yield differences between the two schizophrenic groups, the 10 *S*s rated most reactive (i.e., lowest Elgin ratings) were compared with the 10 rated most process (i.e., highest ratings). Again, no significant differences were found between process and reactive patients in terms of CFF. On the possibility that a relationship existed between a schizophrenic *S*'s flicker rank and his rank on the Elgin Scale, Kendall's tau was computed. The resulting correlation of .048 indicated essentially no relationship between the magnitude of the schizophrenic's Elgin rank (or position on the process-reactive continuum) and his ranked CFF score. The ability of CFF to discriminate brain damaged from all other patients tested was exceptionally good, however. When compared with normals, organic *S*s were successfully discriminated by their flicker threshold 95% of the time since only one organic *S* obtained a higher CFF score than that of the lowest normal.[4] Somewhat less discrimination was present in comparison of process and reactive groups with the organics (90% and 75% respectively).

Archimedes Spiral

With the exception of the 20 brain damaged patients, all of the *S*s (i.e., 20 normals, 20 process, and 20 reactives) obtained perfect scores on the task. Although χ^2 revealed that organics were significantly different from each of the other groups ($p < .01$) the ability of the spiral technique to discriminate brain injured subjects from nonorganic individuals was relatively poor. Of the 20 organics only 9 failed to obtain perfect scores on the task. This resulted in a test discrimination index of 45%. Only 5 of the 9 organics were totally unable to perceive the effect on any of the four trials. No relationship was observed between organics' scores and their neurological diagnoses.

[4] This *S* was a frontal lobe injury.

Discussion

There were no differences in CFF between schizophrenics classified process and reactive. Although the method of classification may be speculated on as one reason for this failure, other studies utilizing the method have differentiated the two groups (Becker, 1955; King, 1958). Present results do not suggest demonstrable cortical defect as measured by CFF in either process or reactive schizophrenia. The fact that the reactive group had a higher CFF than did the normals warrants speculation. Heightened physiological responsiveness of reactive schizophrenics noted in previous studies was attributed by King (1958) to the heightened homeostatic response which reflects the precipitous nature of their adjustment. Other investigations (Ricciuti, 1948; Wiersma, 1906) have indicated that manic patients have higher CFFs than do other groups. Similarity between manic or affective symptoms and characteristics of schizophrenics who might be classified reactive has also been noted in previous studies (Cameron, 1938; Funkenstein, Greenblatt, & Solomon, 1948). Affective components in the clinical picture are among the bases for classification as reactive in the rating scale employed in the present investigation. Therefore, it seems possible that some patients who formerly might have been diagnosed manic-depressive are now being diagnosed schizo-affective or reactive schizophrenic, and these manic features may account for the reactive group's higher than normal CFFs. On the other hand, the significant difference found between normals and reactives may merely indicate a comparison of the extremes in a homogeneous distribution.

A possible alternative explanation of the results might contend that a defect in the experimental design was responsible for lower flicker thresholds in the organics. This position would maintain that since only ascending series were used in the present study it is possible that the organics were responding prior to their subjective impressions of complete fusion and that this premature reaction would also have been obtained in descending

series, thus changing the results. However, since previous investigations (Irvine, 1954; Teuber & Bender, 1949; Werner & Thuma, 1942) utilizing both ascending and descending series have yielded consistently lower thresholds for organics, it would seem more parsimonious to accept a lower CFF threshold on the part of organics, rather than premature reaction, as the explanation for these results.

The present experiment appears to provide additional support for CFF as a good technique in discriminating brain damaged persons with gross destruction from normals and schizophrenics. Increased utilization of the technique in the ordinary hospital setting seems to be well warranted, a possibility recognized by other investigators. Its potential seems especially good when a comparison is made with the relatively poor discriminating ability shown by the usual paper and pencil tests of organicity.

Although the spiral aftereffect differentiated organic *Ss* from all other groups, only 45% of the brain damaged *Ss* were identifiable through failure to perceive the effect. Since only known organic cases were included in the present study it seems unlikely that this technique will be of extensive diagnostic value in a clinical setting where less severe organic cases are to be classified. The fact that none of the schizophrenics of either type failed to perceive the aftereffect does not, at least, support the position that cortical defect is involved in process schizophrenia.

Summary

The present experiment was designed to investigate the possibility of organic involvement, particularly of a cortical nature, in process schizophrenia. Recent studies dichotomizing schizophrenics into process and reactive types have suggested cortical damage in the former group. It was hypothesized, therefore, that patients classified as process schizophrenics would perform similarly to brain damaged individuals on two perceptual tasks, while reactive schizophrenics would more closely resemble normal controls. The two perceptual tasks used were critical flicker frequency and the Archimedes Spiral aftereffect, both of which had proven effective previously in differentiating brain damaged from nonorganic *Ss*. Four groups of 20 *Ss* were given both tasks: reactive and process schizophrenics, organics with known brain damage, and normals. Results of the experiment showed that organic patients were significantly different from all other groups in terms of CFF threshold and ability to perceive the spiral aftereffect. Process and reactive schizophrenics did not differ from each other on either task. Reactive schizophrenics differed from normals on CFF in having higher thresholds. Attempts to account for the findings included the possibility that affective components of reactive schizophrenia contribute to the difference between them and other schizophrenics and normals, and the belief that process schizophrenia, while showing some physiological differences from reactive schizophrenia in other investigations, probably is not related to defects of a cortical nature.

References

Bateman, J., & Papez, J. Significance of the thalamus in psychoses with reference to hallucinations and delusions. *J. clin. exp. Psychopath.*, 1951, **12**, 89–103.

Becker, W. The relation of severity of thinking disorder to the process-reactive concept of schizophrenia. Unpublished doctoral dissertation, Stanford Univer., 1955.

Becker, W. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **53**, 229–236.

Bellak, L. A multiple factor psychosomatic theory of schizophrenia. *Psychiat. Quart.*, 1949, **23**, 738–755.

Bleuler, E. Dementia praecox or the group of schizophrenias. New York: International Univer. Press, 1950.

Brackbill, G., & Fine, H. Schizophrenia and central nervous system pathology. *J. abnorm. soc. Psychol.*, 1956, **3**, 310–313.

Cameron, D. Early schizophrenia. *Amer. J. Psychiat.*, 1938, **95**, 567–582.

Funkenstein, D., Greenblatt, M., & Solomon, H. Autonomic nervous system changes following electric shock treatment. *J. nerv. ment. Dis.*, 1948, **108**, 409–422.

Gallese, A. Spiral aftereffect as a test of organic brain damage. *J. clin. Psychol.*, 1956, **12**, 254–259.

Irvine, R. CFF for paretics and schizophrenics. *J. abnorm. soc. Psychol.*, 1954, **49**, 87–88.

Kallmann, F. The genetics of schizophrenia. New York: Augustin, 1938.

KANTOR, R., WALLNER, J., & WINDER, C. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, **17**, 157–162.

KENNARD, MARGARET, & LEVY, S. The meaning of abnormal electroencephalograms in schizophrenia. *J. nerv. ment. Dis.*, 1952, **116**, 413–423.

KING, G. Differential autonomic responsiveness in the process-reactive classification of schizophrenia. *J. abnorm. soc. Psychol.*, 1958, **56**, 160–164.

NOYES, A. Modern clinical psychiatry. Philadelphia: Saunders, 1953.

PAGE, H., RAKITA, G., KAPLAN, H., & SMITH, N. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, **21**, 89–91.

PRICE, A., & DEABLER, H. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 229–302.

RICCIUTI, H. A comparison of critical flicker frequency in psychotics, psychoneurotics, and normals. *Amer. Psychologist*, 1948, **3**, 276.

ROSS, A., & REITAN, R. Intellectual and affective functions in multiple sclerosis: A quantitative study. *AMA Arch. Neurol. Psychiat.*, 1955, **73**, 663–677.

TEUBER, H., & BENDER, M. Changes in visual perception of flicker, apparent motion, and real motion after cerebral lesion. *Amer. Psychologist*, 1948, **3**, 246–247.

TEUBER, H., & BENDER, M. Alterations in pattern vision following trauma of occipital lobes in man. *J. gen. Psychol.*, 1949, **40**, 37–57.

WERNER, H., & THUMA, B. Critical flicker frequency in children with brain injury. *Amer. J. Psychol.*, 1942, **55**, 394–399.

WIERSMA, E. Die Sekiundarfunktion bei Psychosen. *J. Psychol. Neurol. Lpz.*, 1906, **8**, 1–24.

WINKELMAN, N., & BOOK, M. Observations on the histopathology of schizophrenia: I. The cortex. *Amer. J. Psychiat.*, 1948, **105**, 889–896.

(Received April 16, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

AN ATTEMPT TO EMPLOY A SENTENCE COMPLETION TEST FOR THE DETECTION OF PSYCHIATRIC PATIENTS WITH SUICIDAL IDEAS[1]

HERMAN Y. EFRON
Veterans Administration Hospital, Lyons, New Jersey

Construction of a test suitable for group administration which could detect those who are morbidly preoccupied with depressive thoughts would be extremely useful to ward administrators in hospitals where intensive psychotherapeutic contacts are the exception rather than the rule. This study is an attempt to assess a group screening device (sentence completion test) to identify those patients who may constitute suicidal risks.

The *D* scale of the MMPI (Broida, 1954) and some aspects of responses to the Make-a-Picture-Story (MAPS) (Farberow, 1950) differentiated between suicidal and nonsuicidal patients. However, the few experimental studies with the TAT yielded negative results. Neither responses to Card 3BM (Broida, 1954) nor characteristics of the TAT heroes (Shneidman & Farberow, 1958) distinguished suicidal from control groups. There have been proposed in the literature various Rorschach signs and patterns that can be used to detect potential suicidal behavior (Hertz, 1948, 1949; Piotrowski, 1950; White & Schreiber, 1952); however, other studies generally fail to confirm their validity (Fisher, 1951; Sakheim, 1955). Results indicate that Hertz's (1948) Configuration Technique is probably the most useful of the prognostic scales (Sakheim, 1955), but even her technique failed when used with schizophrenic patients (Fisher, 1951). Even if such a test as the Rorschach were capable of differentiating between potential suicidal patients and nonsuicidal patients, the time involved in arriving at these decisions would preclude their general use. The number of psychologists needed to service a psychiatric hospital would be far in excess of current staffing patterns even though periodic retesting was restricted to those patients whose case histories warrant some suspicion of self-destructive behavior. A self-administering test which could be repeated at specified intervals, tapping both attitudes within and outside the patient's awareness, seemed to be the most desirable instrument. It was felt that the sentence completion test might be such an instrument.

The most difficult problem in designing an experiment to test the efficiency of an instrument to predict suicide is that of an appropriate criterion. Because of the low incidence of suicide it is not practicable to rely upon this as a criterion measure. Rosen (1954) has clearly analyzed the inherent problems in predicting an event whose occurrence among psychiatric patients (as judged from statistics covering 1936 to 1949 at a large state hospital) is roughly 3 per 1,000 admissions[2] (Levy & Southcombe, 1953). In his article, Rosen (1954) demonstrates that if a test were

[1] This study was an individual hospital project of the Veterans Administration Psychiatric Evaluation Project, Richard L. Jenkins, Director. This paper was reviewed in the Veterans Administration and is published with the approval of the Chief Medical Director. The statements and conclusions of the author are the result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration. The author wishes to express his gratitude to Marcus P. Rosenblum for his constructive suggestions, and to Leon Cohen, Eric Layne, David Levine, and John Seymour for their cooperation in rating the sentence completions.

[2] Using the same data, Rosen (1954) incorrectly arrived at .0032 as a rough index of suicide among the psychiatric population. Actually, since there are about three to four times as many resident patients per year as admissions per year, a closer approximation to the actual suicide rate would be one third to one fourth of .0032.

devised that could correctly identify whether a patient was suicidal or nonsuicidal 75% of the time, the number of false positives (i.e., nonsuicidal patients being identified as suicidal) in a patient population of 12,000 would be close to 3,000. Recognizing this severe limitation it was still felt that if a group screening device that would approach 75% accuracy could be developed it would serve to alert the hospital personnel to investigate more thoroughly all patients designated as "suicidal." In a hospital of 2,000 this would mean more vigilant study of about 500 patients, an aim that might be feasible. Rosen (1954) emphasizes the need for researchers to formulate more precise definitions of "suicidal" patients. Studies by Farberow (1950) and Rosen, Hales, and Simon (1954) found marked group differences between the MMPI scores of patients who attempted suicide and those who threatened suicide. The group that made threats gave significantly higher mean scores on most all scales. Both studies indicate that as a group those who made suicidal threats were more severely disturbed (at the time of testing, which was after those in the attempted suicide group had executed their attempts). The present study deals with patients who were known to have expressed conspicuous suicidal thoughts, marked nihilistic ideas, or severe guilt with concern over punishment or torture. A basic assumption is that patients with such morbid ruminations are greater suicidal risks than those who do not have such preoccupations. The only evidence that can be cited to support the assumption bears indirectly upon it and is possibly irrelevant, viz., Shneidman and Farberow's (1956) report that almost all of the people they studied who did commit suicide had given some earlier warning of their intentions. The hypothesis of this study is that psychologists can, on the basis of clinical judgment, differentiate between the sentence completions of patients with suicidal ruminations and those without such ruminations.

Method

Test Instrument

Some incomplete sentences were purloined from existing tests and still others were constructed until the finished test consisting of 54 items was evolved.[3] Almost a third of the items appear to be weighted toward eliciting depressive thema. Included in this group were such items as "If there were no way out I would—," "My greatest fault is—," and "Bob's defeat made him—."

Procedure

Subjects (*Ss*) were tested in small groups of 10 to 18 patients were given one hour in which to complete the test. Throughout the testing session the examiner encouraged those patients who seemed to be having difficulty in completing the sentences. When more than about half the items were omitted the patient was asked to write the reason for his difficulty in completing the sentences. In a few instances the patients even refused to do this, and the examiner wrote the reason the patient stated verbally.

Phase I. Of the 104 obtained completions 54 were given to four staff psychologists. They were told that 18 were done by suicidal patients, 9 by assaultive patients, 27 by patients who were neither suicidal nor assaultive, and they were then asked to distribute them into their proper categories. After every judgment they were requested to state the degree of confidence they had in their judgment. Coding for these statements was: 1. I'm in doubt whether this is correct. 2. I feel fairly confident that this correct. 3. I feel confident that this is correct. The only identifying data were the patient's code number and age.

Phase II. From the 54 completions used in Phase I, all those with less than 25 responses were eliminated. The same four psychologists were asked to redistribute the tests which now represented 10 suicidal patients, 3 assaultive patients, and 25 nonsuicidal and nonassaultive patients. They were again asked to rate their confidence in each decision.

Phase III. In order to obtain an estimate of the frequency of false positives, the remaining 50 completions (which were all done by patients who were nonsuicidal and nonassaultive) were given to three of the four psychologists with the following instructions:

> The following sentence completion tests were obtained from patients who may have been on suicidal observation status, on assaultive status, on *suicidal and assaultive* status, or on none of the above. On the accompanying Summary Sheet, will you please distribute the code numbers of the patients into their appropriate categories. Remember the frequency in any of the above categories may be zero.

This time there was also a "cannot rate" category in the event they found it impossible to reach a decision. Again they were requested to state the degree of confidence they had in each judgment.

Subjects. Patients were selected from the acute intensive treatment service of a large Veterans Ad-

[3] A copy of the incomplete sentences can be obtained from the author.

TABLE 1

DESCRIPTION OF SAMPLE

Demographic Variable	Phase I		Phase II
	Suicidal	Nonsuicidal	Nonsuicidal
N	18	36	50
Diagnosis			
Schizophrenic	18	33	44
Manic-depressive	0	2	3
Chronic brain syndrome	0	1	1
Other	0	0	2
Current hospitalization			
Less than 6 months	7	15	16
Six months to 2 years	2	13	16
More than 2 years	9	8	18
All NP hospitalizations			
Less than 6 months	4	1	2
Six months to 2 years	2	12	18
More than 2 years	12	23	30
Age			
Mean	35	37	36
SD	5	11	9

ministration hospital. For a patient to have been considered suicidal he must have expressed serious self-destructive thoughts, marked nihilistic ideas, or severe guilt with concern over punishment and torture to the ward physician, nurse, or aide within the week prior to testing. To have been considered assaultive [4] the patient, in the estimation of the ward physician, had to present a high risk of acting out aggressively. Those patients who, for at least two months prior to testing, had not expressed any self-destructive or depressive ideas and had no known history of suicidal attempt were considered nonsuicidal. Nonassaultive patients were those whom the ward physician had no reason to suspect would act out aggressively. Of the 104 *Ss*, 18 were suicidal, 9 were assaultive, and 77 were both nonsuicidal and nonassaultive. Of the latter group, 33 were privileged (i.e., had freedom of the hospital grounds) and 44 were nonprivileged. Table 1 contains a further description of the groups.

RESULTS AND DISCUSSION

The number of correctly assigned suicidal patients was compared with the chance expectancy of such an occurrence as computed from the hypergeometric function. For Phase I only one of the four psychologists' distributions warranted assuming that other than chance factors were operating. The probabilities obtained were .82, .38, .06, and .02. When the tests which contained less than 25 responses were discounted (Phase II) none departed significantly from chance expectation ($P = .97, .53, .53, .06$). The over-all percentage of correct identification of patients with suicidal ruminations was 43% for Phase I and 30% for Phase II. Since in Phase III the true distribution was not disclosed to the psychologists, their ratings provide us with an index of the false positives we might expect. The mean percentage of false positives was 33%.[5] Clearly these results indicate that the test as used is woefully inadequate to provide practical value in the detection of suicidal patients.

A comparison was made between the num-

[4] Assaultive patients were included in the study in order to approximate the clinical situation and thereby encourage each rater to evaluate the completions more critically. Also, it was intended to prevent spurious positive findings that may have resulted by virtue of the fact that more severely disturbed patients have a greater likelihood of being rated suicidal.

[5] It is possible that the set established in Phases I and II by virtue of the raters being told the number of suicidal patients to expect resulted in increasing the probability of false positives. However, for one rater Phase III preceeded the others. His percentage of false positives was 30%.

ber of agreements among psychologists with the number to be expected if chance were the only factor. The expectancy derived from probability theory was compared with the obtained results by means of chi square. For the purpose of these comparisons all the distributions were dichotomized into the categories "suicidal" and "others." The chi squares for Phase I and Phase II were significant beyond the .01 level of significance, and for Phase III beyond the .05 level. When the assaultive patients were grouped with the suicidal patients the chi squares for Phase I and II were significant beyond the .001 level and for Phase III beyond the .01 level. This latter finding coupled with the previous observation that the psychologists did somewhat better when records with less than 25 responses were included tends to indicate that their ratings were more in terms of severity of illness than focus upon specific traits that may be associated with suicide. Further evidence to support this formulation is available from the comparison of the number of privileged and nonprivileged patients rated as suicidal. Although both these groups were actually nonassaultive and nonsuicidal, the likelihood of a nonprivileged patient being rated as suicidal was 60% greater. Since suicidal ruminations may be the result of varied and wholly different intrapsychic processes, our findings are not surprising.

In analyzing the relationship between the degree of confidence and correctness of decision it was found that with the exception of the one psychologist who did better than chance in Phase I there was no marked increased accuracy when a decision was made with confidence. The exceptional rater was correct in 11 of his 14 high confidence decisions and only 4 of his 14 low confidence decisions. Such a difference would occur on a chance basis less than one time in a hundred (chi square with Yates correction). Using only the high confidence ratings of this last rater he was able to place correctly 79% of the suicidal patients. It would therefore seem that such a test as the one described may, in the hands of a proficient rater, have some usefulness in detecting suicides. Further research focused upon identifying proficient raters seems warranted.

SUMMARY

One hundred and four psychiatric patients were given a sentence completion test constructed to elicit suicidal ideation. Of the 104 subjects, 18 were known to have been preoccupied with suicidal ruminations. In one part of the experiment, four clinical psychologists categorized sentence completions as coming from suicidal, assaultive, and nonassaultive patients after having been given the frequencies in each category. In another part of the experiment, three psychologists were given different sentence completions and asked to categorize them without being told the actual distribution.

The over-all percentage of correct identification of patients with suicidal ruminations varied from 30% to 43%. The mean percentage of false positives was 33%. Although the psychologists' predictions showed no greater relationship to the criterion measure than chance expectancy, they agreed with each other more often than would be expected if chance were the only factor. This, coupled with other evidence, was taken to indicate that the psychologists' ratings were primarily in terms of severity of illness rather than focusing on specific suicidal traits. As a group, the raters' performance was not significantly better when they had high confidence in their decisions. However, the high confidence ratings of one psychologist correctly placed 79% of the suicidal patients. This was taken to indicate that there was a possibility of identifying proficient raters who could use this test successfully.

REFERENCES

BROIDA, D. C. An investigation of certain psychodiagnostic indications of suicidal tendencies and depression in mental hospital patients. *Psychiat. Quart.*, 1954, **28**, 453–464.

FARBEROW, N. L. Personality patterns of suicidal mental hospital patients. *Genet. Psychol. Monogr.*, 1950, **42**, 3–79.

FISHER, S. The value of the Rorschach for detecting suicidal trends. *J. proj. Tech.*, 1951, **15**, 250–254.

HERTZ, MARGUERITE R. Suicidal configurations in Rorschach records. *Rorschach Res. Exch.*, 1948, **12**, 3–58.

HERTZ, MARGUERITE R. Further study of "suicidal" configurations in Rorschach records. *Rorschach Res. Exch.*, 1949, **13**, 44–73.

LEVY, S., & SOUTHCOMBE, R. H. Suicide in a state hospital for the mentally ill. *J. nerv. ment. Dis.*, 1953, **117**, 504–514.

PIOTROWSKI, Z. *A Rorschach Compendium*. Utica: State Hospital Press, 1950.

ROSEN, A. Detection of suicidal patients: An example of some limitations in the prediction of infrequent events. *J. consult. Psychol.*, 1954, **18**, 397–403.

ROSEN, A., HALES, W. M., & SIMON, W. Classification of "suicidal" patients. *J. consult. Psychol.*, 1954, **18**, 359–362.

SAKHEIM, G. A. Suicidal responses on the Rorschach test: A validation study. *J. nerv. ment. Dis.*, 1955, **122**, 332–344.

SHNEIDMAN, E. S., & FARBEROW, N. L. Clues to suicide. *U. S. Public Hlth. Rep.*, 1956, **71**, 109–114.

SHNEIDMAN, E. S., & FARBEROW, N. L. TAT heroes of suicidal and non-suicidal subjects. *J. proj. Tech.*, 1958, **22**, 211–228.

WHITE, M. A., & SCHREIBER, H. Diagnosing suicidal risks on the Rorschach. *Psychiat. Quart. Suppl.*, 1952, **26**, 161–189.

(Received April 17, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

FACTOR ANALYSES OF "ANXIETY" AND "NEUROTICISM" INVENTORIES

A. W. BENDIG
University of Pittsburgh

In recent years a number of different questionnaires purporting to measure "anxiety" (Cattell, 1957b; Taylor, 1953) or "neuroticism" (Cattell, 1955; Eysenck, 1956a; Winne, 1951) have been independently developed by different methodologies, and subsequent studies have shown relatively high correlations among these scales (Bendig, 1957; Holtzman, Calvin, & Bitterman, 1952; Windle, 1955). These correlations suggest that the Eysenck, Winne, and Taylor scales may all be measuring the same objective personality dimension. However, several attacks have been launched against the undimensionality of these scales. Eysenck (1956b, 1957) argues that such questionnaires are confounding the independent dimensions of neuroticism and extraversion and his hypothesis has received some experimental verification (Bendig, 1957, 1958; Frey & Becker, 1958). A second objection to such scales has been that of Edwards (1957) who suggests that much of the variance of such scales can be attributed to social desirability or the tendency of *Ss* to give socially desirable responses to personality items regardless of their content and the large correlations between Edwards' Social Desirability Scale (SDS) and Taylor's MAS (Edwards, 1957, p. 33) and Cattell's Anxiety scale (Bendig, 1959b) have supported this hypothesis. However, 22 of the 39 items on Edwards' scale overlap with Taylor's MAS; and when only the 17 non-MAS items are scored on the Social Desirability Scale, Edwards found the MAS correlation substantially reduced (Edwards, 1957, p. 88), and Bendig (1959b) similarly reported that the Cattell Anxiety score correlation decreased. These findings suggest that the Edwards SDS is also not undimensional, but measures a combination of social desirability, anxiety, and/or neuroticism factors. Parenthetically it might be noted that the SDS shows a low positive correlation with the MMPI Lie scale (Edwards, 1957, p. 44), suggesting that the SDS scores may also involve a "lie" component.

At a theoretical level, Spence (1958) argues for the construct validity of Taylor's MAS as an operational definition of "manifest anxiety," now renamed "emotionality" or "emotional reactivity," while Eysenck (1956b; 1957), as noted above, regards anxiety as a combination of intraversion and neuroticism. Although Spence may be regarded as a monist in this area of personality theorizing, and Eysenck as a dualist, Cattell (1957a) is a trinitarian since he regards anxiety, neuroticism, and extraversion as three relatively independent personality dimensions and has published inventories for the measurement of both neuroticism and anxiety (Cattell, 1955, 1957b). So we are left with one, two, or three personality factors with varying identifications depending on the theoretical orientation of the scientist.

The present research was designed to investigate the number of orthogonal factors present in a battery of independent questionnaires purportedly measuring the above noted personality dimensions and to clarify the factorial composition of each of these commonly used inventories. It was assumed that the intercorrelations among the anxiety and neuroticism scores might be influenced by other contaminating factors, such as extraversion, social desirability, falsification, and omnipresent sex differences, and the inclusion of marker variables for these hypothetical factors would help to purify the factors obtained. The central question of the study was

whether, when the influence of other factors was isolated by these marker variables, a single general factor of "emotionality" (Spence, 1958) or separate group factors of "anxiety" and "neuroticism" (Cattell, 1957b) would appear necessary to account for the interrelationships among the scales.

Study I

Procedure

Scales. The scales included in this study were assembled into a single booklet containing 230 items from which 12 scores were obtained. Only 10 scores were used with each of two groups of *Ss*, since four of the scores were not experimentally independent (see below).

1. Manifest Anxiety Scale (Taylor, 1953) included only the 50 dichotomous-response items that are actually scored for this scale (McCreary & Bendig, 1954).[1]

2. Winne's (1951) Neuroticism scale: Since 8 of the 30 dichotomous-response MMPI items on this scale are included in the Manifest Anxiety scale, only the 22 experimentally independent items were used.[2]

3. Edwards' (1957) Social Desirability Scale: Only the 17 dichotomous-response MMPI items that are not included in Taylor's MAS were scored for this scale.[3]

4. Eysenck's (1956a) Extraversion scale, which is included in the Maudsley Personality Inventory, contains 24 items.

5. Eysenck's (1956a) Neuroticism scale is the second 24-item scale included in the MPI. Dichotomous item responses (Yes-No) were used with items on both of the Eysenck scales (Bendig, 1959a).

6. MMPI Lie scale (Hathaway & McKinley, 1951): Since two of the standard 15 items appeared inappropriate for our college *Ss*, only 13 dichotomous-response items were used.[4]

7. Cattell's (1957b) Covert Anxiety scale included the first 20 "subtle" trichotomous-response (true-?-false) items of the IPAT Anxiety scale.

8. Cattell's (1957b) Overt Anxiety scale included the last "overt, symptomatic" items of the IPAT Anxiety scale.

9. Cattell's (1956b, p. 7) Difference score was the score for Overt Anxiety minus the score for Covert Anxiety. Cattell suggests that the difference between the Covert and Overt scores "is intended to give a measure of the degree to which the patient is or is not conscious of his anxiety, and perhaps also of any wish consciously to emphasize it."

10. Cattell's Total Anxiety score was the sum of the Covert Anxiety and Overt Anxiety scores. Since these four scores derived from the IPAT Anxiety scale are not all experimentally independent, the Covert Anxiety and Overt Anxiety scores (which are independent) were included in the analysis of the data from one second group of *Ss* (Group A), while the Difference and Total Anxiety scores (which also are independent) were used in the analysis of the data from the second group of *Ss* (Group B).

11. Cattell's Neuroticism scale contained the 34 trichotomous-response items scored in the IPAT Neurotic Personality Factor questionnaire (Cattell, 1955). The items were reversed in scoring (high score representing high neuroticism) for consistency in scoring direction with other similar scales included in this study.

12. Cattell's (1956b) Motivational Distortion scale contains the remaining six trichotomous-response items in the IPAT NPF inventory (Cattell, 1955). Cattell's description of the construction of this scale suggests that it is measuring a factor similar to that assessed by the MMPI Lie scale.

In the above list, Variables 3, 4, 6, and 12 were included as marker variables for possible contaminating factors, while Variables 1, 2, 5, 7, 8, 9, 10, and 11 provided the test of the single Emotionality factor versus the double Anxiety-Neuroticism factor hypotheses.

Subjects. The scales were administered during a regular class session to 11 sections of introductory psychology, with a total of 425 *Ss* (261 men and 164 women) being tested. The *Ss* are primarily freshman and sophomore students in the School of the Liberal Arts with smaller proportions of juniors and seniors and of students enrolled in the School of Business Administration and the School of Engineering and Mines. After scoring, the answer sheets were alphabetized by *Ss*' names and the first 118 male *Ss* and the first 82 female *Ss* selected as Group A. In a similar manner the next 118 male and 82 female *Ss* were selected as Group B. The mean age of the 400 *Ss* was 20.8 years, with a standard deviation of 4.5 years. As noted above, the only procedural difference in analyzing the data from each group was that Cattell's Covert Anxiety and Overt Anxiety were included as score variables for Group A, while Cattell's Difference and Total Anxiety scores were used in the analysis of the data from Group B. *Ss*' sex (male scored 1, female scored 0) was included in each analysis as the eleventh variable.

Method. The 10 questionnaire variables within each group of *Ss* were intercorrelated by the usual product-moment method, and these variables were correlated with the sex dichotomy by the point biserial formula. Each matrix of 55 intercorrelations was factor analyzed by the complete centroid method, and four factors were extracted. Each of the succes-

[1] The MMPI booklet numbers can be found in Windle (1955).

[2] The MMPI booklet numbers and direction of scoring of these items are: 5, 29, 41, 44, 47, 72, 76, 108, 114, 159, 189, and 236 scored True; and 2, 3, 9, 46, 51, 68, 103, 175, 178, and 208 scored False.

[3] The MMPI booklet numbers and direction of scoring of these items are: 54, 169, and 257 scored True; 40, 42, 138, 148, 156, 171, 218, 245, 247, 252, 267, 269, 286, and 383 scored False.

[4] The MMPI booklet numbers are: 15, 30, 45, 60, 75, 90, 105, 120, 150, 165, 195, 225, and 285 scored False.

TABLE 1

ROTATED ORTHOGONAL FACTOR LOADINGS (DECIMAL POINTS OMITTED) OF INVENTORY SCORES

Scales	Group A				Group B			
	Em	F	Sx	h^2	Em	F	Sx	h^2
Taylor Manifest Anxiety Scale	83	−19	21	77	82	−32	12	79
Winne Neuroticism (Abbreviated)	61	−05	05	37	62	−35	−05	51
Edwards Soc. Desir. (Abbreviated)	−57	39	17	51	−54	46	14	53
Eysenck MPI Extraversion scale	−36	−04	−08	14	−46	−05	−03	22
Eysenck MPI Neuroticism scale	75	−33	12	68	67	−50	18	73
MMPI Lie scale (Abbreviated)	−07	66	05	44	−08	67	02	45
Cattell Covert Anx. subscale	50	−64	00	66				
Cattell Overt Anx. subscale	60	−37	40	66				
Cattell Difference score					47	09	37	37
Cattell Total Anxiety scale					61	−66	21	85
Cattell NPF Neuroticism scale	34	−09	66	56	42	−37	55	61
Cattell NPF Motiv. Dist.	−05	62	−12	40	−05	64	05	41
Sex (Female)	−02	03	56	31	−04	08	54	30

sive factors was tested for statistical significance using both Tucker's and Humphrey's criteria (Fruchter, 1954, pp. 77–80). The significant factors in each matrix were analytically rotated to orthogonal simple structure using the normalized varimax method (Kaiser, 1958).[5]

Results

The first four centroid factors from Groups A and B accounted for 53.2 and 54.8% of the variance of the 11 variables within each group of *Ss*, with the Group A factors accounting for 34.6, 9.1, 6.3, and 3.2 of the variance, while the similar percentages from Group B were 38.6, 8.7, 5.2, and 2.3. The application of Humphrey's rule in testing the significance of the factors indicated that the first three factors from Group A and the first two factors from Group B were significant at the .05 level or better. Using Tucker's procedure the first three factors were significant within both groups. Consequently, only three factors were retained for the normalized varimax rotation, with the fourth factor being discarded as not being significant under either criteria.[6]

The loadings of each variable resulting from the normalized varimax rotation procedure are given in Table 1. For convenience and consistency, the signs of the loadings for the Sex variable were reversed in preparing Table 1 as is indicated in the table. The three factors are almost identical between both Groups A and B, with indices of factor similarity (sum of the cross-products of loadings of similar factors divided by the geometric mean of the sums of squares of the loadings) for the nine variables common to both Groups A and B being .99, .94, and .96 for rotated Factors Em, F, and Sx. The only evident sampling difference between the independent groups of *Ss* is that the loadings of the scores on Factor F tend to be higher within Group B than they are within Group A. The relative rank orders of the Factor F loadings are almost identical within both groups (rho = .92), but F accounts for only 15% of the interscale variance within Group A, while it accounts for 19% within Group B.

Factor Em is an Emotionality factor with zero loadings only on the MMPI Lie, Cattell Motivational Distortion, and Sex variables. Taylor's MAS is the best single measure of Factor Em, with Eysenck's Neuroticism, Winne's Neuroticism, and Cattell's Total Anxiety scores also having large loadings on this factor. This Emotionality factor, as de-

[5] The author's appreciation is extended to Gary Lotto and William B. Kehl of the University's Computing and Data Processing Center for providing assistance and facilities in the factor analytic phase of this study.

[6] Tables giving the intercorrelations among the variables within each group of *Ss*, the tests of significance of the unrotated centroid factors, and centroid loadings and transformation matrices have been deposited with the American Documentation Institute. Order Document No. 6513, remitting $1.25 for microfilm or $1.25 for photocopies.

fined by this battery of inventories, permeates both the Anxiety and Neuroticism scales equally strongly and shows a moderate negative loading (−.36 and −.46) on the Extraversion scale. Factor F is a Falsification factor with high loadings on the MMPI Lie and Motivational Distortion scales. Since neither of these two variables was loaded on the other two factors, they can be used to define and identify this factor. Most of the scales with high Emotionality loadings also have moderate negative loadings on this F factor with Cattell's Covert Anxiety and Total Anxiety scales showing the largest loadings. Eysenck's Extraversion scale and the Sex variable have zero loadings on Factor F. Factor Sx appears to be a pure Sex factor with the reversed direction of scoring of this variable identifying the positive pole as the "femininity" direction. Both Cattell's Overt Anxiety and Neuroticism scales have large positive loadings on the Sx factor as does Cattell's Difference score (which is probably attributable to the inclusion of Covert Anxiety in the Difference score).

Considering the factor composition of the inventory scores included in this analysis, the Social Desirability, Eysenck Neuroticism, Covert Anxiety, Total Anxiety, and probably the Manifest Anxiety and Winne Neuroticism scales are factorily complex by being loaded on both Factors Em and F. Overt Anxiety and probably Cattell's Neuroticism scales are loaded with all three factors, while Cattell's Difference score is loaded with both Factors Em and Sx. Eysenck's Extraversion scale shows a moderate loading on Factor Em, and the MMPI Lie and Motivational Distortion scales are the best measures of Factor F.

Discussion

The absence of a significant unrotated fourth factor within either matrix of inventory intercorrelations and the clear identification of the analytically rotated three significant factors as Emotionality, Falsification, and Sex factors indicate that the Anxiety and Neuroticism dimensions measured by the scales developed by different authors and different methodologies can be assumed to be operationally identical within the limits of this study. This is not to deny that some of these scales may be also measuring factors that had no opportunity to appear in these analyses, but the communalities of several scales, notably Taylor's MAS, Eysenck's Neuroticism, Cattell's Covert Anxiety, Overt Anxiety, Total Anxiety, and Difference scores, are close to what might be assumed to be the reliabilities of these scales, leaving little reliable scale variance to be attributed to hypothetical factors not appearing in these analyses.

The Emotionality factor defined by these inventories includes an introversion component as has been suggested by Eysenck (1957). His further hypothesis that this component results from the confounding of an orthogonal Extraversion-Introversion (EI) factor with Em could not be tested with the present data since only one "pure" marker variable for the hypothetical EI factor had been included in the analyses. Inclusion of additional measures of EI might permit a separation of Emotionality and Extraversion factors that may have been confounded in the present analyses and allow the assessment of the relative proportions of each factor in these Anxiety and Neuroticism scales. The presence of a Falsification or Honesty factor in these inventory scores, as was suggested by Edwards (1957) among others, was also strongly confirmed, but it is evident that Edwards' Social Desirability Scale is not the best measure of this factor. The factor loadings of the abbreviated Social Desirability Scale appear quite similar to those of the Anxiety and Neuroticism scales and indicate that the SDS scores are better measures of Factor Em than they are of Factor F. Although Eysenck (1956a) was moderately successful in developing an Emotionality scale that is independent of Extraversion (Bendig, 1957), his scale is also negatively loaded with the Falsification factor.

The study reported below is a further extension of the first study with additional marker variables for the Em and TH factors. It was hoped that these additional scales would clarify the relation of the scales to the Em, EI, F, and Sx factors and would again test for separate Anxiety and Neuroticism factors.

Study II

Procedure

Scales. Eleven independent personality scales were included, along with standard instructions, in a booklet to be distributed to the *Ss* with separate IBM answer sheets being used to collect the responses to the items.

Maudsley Personality Inventory (MPI) scores: The most recent form of the MPI, described in detail by Jensen (1958), consists of 80 trichotomous-response (Yes-?-No) items including 24 Extraversion, 24 Neuroticism, 20 Lie, and 12 Buffer items. Eysenck (1956a) describes the 12 Buffer items as 7 nonscored Extraversion and 5 nonscored Neuroticism items to be used for research purposes. He also indicates which of the 24 items on either the E or the N scales can be split into subscales containing 12 items each for establishing the split-half reliability of the E and N scales. Much of the content of the Lie items is similar to that in other Lie scales such as the MMPI scale described below. Although the original MPI uses the trichotomous-response method, it has been shown that the reliability and intercorrelation of the scales are not affected if a simpler dichotomous-response (Yes-No) method is used (Bendig, 1959a).

For purposes of the present research, the items of the MPI, using the dichotomous-response format, were grouped into six independent scales to obtain separate scores. MPI-E1 and MPI-E2 each contained 12 of the original 24 Extraversion items with the assignment of items to subscales following the split suggested by Eysenck (1956a). MPI-E3 included the 7 Buffer MPI Extraversion items. MPI-N1 contained the 24 items in the original Neuroticism scale, while MPI-N2 was scored for the 5 Buffer MPI Neuroticism items. MPI-L scores were obtained by scoring 10 of the 20 MPI Lie items, omitting 10 Lie items that either overlapped in item content some of the other Lie scales used in this study or appeared not to be items which would be relevant for our *Ss*.[7]

Some of the scales included in the previous study were repeated in this analysis. Taylor's 50-item Manifest Anxiety Scale (MAS) was included along with the abbreviated 22-item version of Winne's Neuroticism (Winne N) scale and the 13-item form of the MMPI Lie scale. The 34-item Neuroticism (NPF-N) and 6-item Motivational Distortion (NPF-MD) scales included in Cattell's Neurotic Personality Factor Questionnaire were also repeated in this study. As in the previous study, the NPF Neuroticism items were reversed in scoring compared to the item key given in the NPF manual so that high scores here represent high Neuroticism.

In summary, the MPI-N1, MAS, Winne N, MMPI Lie, NPF-N, and NPF-MD scales are identical with those used previously, while the MPI-E1 and MPI-E2 combined are equivalent to the previous Extraversion scale. MPI-E3, MPI-N2, and MPI-L were used for the first time in this study.

Subjects. The scales were administered to a total of 263 students (197 men and 66 women) enrolled in eight sections of a one-semester course in introductory psychology (Group C). The *Ss* were primarily freshman and sophomore students and are similar to those used in the previous study with the exception that a smaller percentage of women *Ss* was enrolled during this (spring) semester. The scales were administered in class during the second week of the semester by each class instructor, and this group of *Ss* represents the total enrollment in the course during this semester with the exception of a few *Ss* who were absent from class the day that the scales were administered.

Method. The 11 scale scores were intercorrelated by the usual product-moment method, and each score was correlated with the sex dichotomy by the point biserial formula. Six factors were extracted from the 12-variable matrix of correlations by the complete centroid method and each factor was tested for statistical significance using both Humphrey's and Tucker's criterion. The significant factors were rotated to orthogonal simple structure using the analytic normalized varimax criterion.

Results

Using Humphrey's test of the significance of factors, the first three centroid factors were significant at the .01 level and the fourth factor at the .05 level ($z = 2.20$), while the fifth and sixth factors did not approach significance ($z = .82$ and .53). The first five factors were significant using Tucker's phi criterion, while the sixth factor again was not significant. Including the fifth factor in the rotation process resulted in a fifth rotated factor with all loadings less than .30 (the largest loadings were −.28 for NPF-N and .22 for MMPI Lie). Consequently, the fifth factor was omitted and only the first four factors included in the normalized varimax rotation procedure. The four unrotated centroid factors accounted for 53.8% of the total variance with the omitted fifth factor accounting for only an additional 2.0%. The residual correlations after the extraction of four factors ranged in absolute

[7] The content of the MPI items and item scoring direction can be found in Jensen (1958, pp. 316–318). The numbers of the items included in this study are: MPI-E1, 1, 3, 6, 13, 14, 20, 31, 56, 57, 66, 69, 75; MPI-E2, 2, 9, 23, 37, 41, 44, 46, 47, 51, 61, 73, 79; MPI-E3, 7, 21, 29, 34, 60, 70, 80 (all scored for Yes responses); MPI-N2, 22, 30, 53, 74, 78 (all scored Yes); and MPI-L, 5, 12, 19, 33, 36, 40, 43, 49, 58, 62. The MPI-N1 scale included the 24 items listed by Jensen for the MPI-N scale. Lie items 8, 15, 24, 28, 45, 52, 55, 67, 72, and 77 were omitted from the form given to the *Ss*.

TABLE 2

ROTATED ORTHOGONAL FACTOR LOADINGS (DECIMAL POINTS OMITTED) OF INVENTORY SCORES (Group C)

Scales	Em	EI	F	Sx	h^2
Eysenck MPI-N1	78	−07	−25	14	70
Eysenck MPI-N2	64	−01	−13	10	43
Taylor MAS	85	−23	−12	07	80
Winne N	69	−23	−04	04	53
Cattell NPF-N	46	04	−13	52	50
Eysenck MPI-E1	−22	78	01	−10	66
Eysenck MPI-E2	−10	79	−02	11	64
Eysenck MPI-E3	−05	79	08	02	64
Eysenck MPI-L	02	04	62	06	39
MMPI Lie	−27	−10	56	10	40
Cattell MPF-MD	−22	11	55	−11	38
Sex (Female)	05	01	09	59	36

values from .00 to .08, with the median being .015.

The rotated factor loadings show a good approximation to simple structure, with 45.8% of the loadings having absolute magnitudes of .10 or less, 60.4% being .20 or less, and 27.1% of the loadings being significant (.30 or greater). Only NPF-N had significant loadings on two factors with the other 11 variables having loadings above .30 on one factor. The factor loadings and communalities of the scales can be found in Table 2. The first factor is obviously the Emotionality (Em) factor with high loadings on all of the Anxiety and Neuroticism scales, while the orthogonal second factor is an Extraversion-Introversion (EI) dimension which loads highly on the three Extraversion subscales. The third rotated factor again is a Falsification (F) factor with high loadings on the three Lie scales and the fourth factor is a Sex (Sx) factor that loads only on the NPF-N and Sex (female) variables. MPI-N1 has a small negative loading on F, while the MAS and Winne N variables have small negative loadings on EI. MMPI Lie and NPF-MD scales have small negative loadings on Em in addition to their large loadings on F with MPI-L loading only on F. NPF-N loads on both Em and Sx with its loading on Em (.46) being slightly smaller than its loading on Sx (.52).

Discussion

The results of this factor analysis have confirmed and clarified the findings of the previous analyses. The presence of an orthogonal EI factor is confirmed and has been separated from the Em factor, while the previous analyses confounded these two factors. Again only one major Em factor was found to load the Anxiety and Neuroticism scales, which confirms the previous evidence that Anxiety and Neuroticism factors cannot be demonstrated as separate factors in factor analyses of questionnaire data. The identification of the factor was strengthened and more clearly separated from the EI factor.

Using the squares of the factor loadings in Table 2 as percentages of scale variance, we can quantitatively estimate factor components in the scales included in this research. Of the Anxiety and Neuroticism scales, the MAS is the most valid measure of Em (72%), but is contaminated by a small loading on EI (5%). Eysenck's MPI-N1 score is also a good measure of Em (61%) and is free of EI, but is slightly loaded on F (6%). Winne's Neuroticism scale shows a lower weighting on Em (48%) and, like the MAS, has an EI component (5%). Cattell's NPF Neuroticism scale is a better measure of Sex differences (Sx) than it is of Emotionality (21% vs. 27%) which agrees with the previous analyses. The greater saturation of the MAS score with Em may be due to the greater length of this scale (50 items) compared to the other scales, assuming that a longer scale would have a greater proportion of reliable variance and, possibly, a larger proportion of common factor variance.

Approximately 62% of Eysenck's Extraversion scale variance is accounted for by the EI factor and 3% is attributable to Em. All three of the Lie scales (MPI-L, MMPI Lie, and NPF-MD) load moderately on the F factor (38, 31, and 30%), but MMPI Lie and NPF-MD also have Em components (7 and 5%). MPI-L appears to be the purest measure of this F factor with negligible loadings on Em, EI, and Sx. It is interesting to note that the three scales composed of MMPI items (MAS, Winne, and MMPI Lie) all have both Em and EI components.

In using these scales for further factor analytic studies, MAS, MPI-N1, and MPI-N2 appear to be the best "markers" for the Em factor and Cattell's and Winne's Neuroticism scales can be eliminated as being too contaminated by extraneous factor variance. However, the Winne scale might be improved and purified through factor analysis. The MPI Extraversion scores can be used to define the EI factor, but should be improved to remove the F component. MPI-L can be used as a measure of F, but its F saturation should be increased, possibly by lengthening the scale, and at least one other F marker scale will have to be developed that is free of the Em factor.

Summary

Various inventories measuring Anxiety, Neuroticism, Extraversion, and Falsification variables were administered to two groups of college students, each group containing 200 *Ss*, and replicated factor analyses of the intercorrelations among 11 scores within each group were performed. Three significant factors were found within each matrix of correlations and the factors were rotated to orthogonal simple structure by the normalized varimax method. The rotated factors were identified as Emotionality, Falsification, and Sex with all of the Anxiety and Neuroticism scores having appreciable loadings on both of the first two factors. It was suggested that Anxiety and Neuroticism are both manifestations of a more general Emotionality factor and are not separate dimensions within commonly used inventories, and that the analyses may have confounded two independent factor dimensions, Emotionality and Extraversion-Introversion, within the Emotionality factor found in these two samples.

To test the above findings, a second study was conducted in which 11 Anxiety, Neuroticism, Extraversion, and Lie scales overlapping with the first battery were administered to 263 students and a centroid factor analysis performed on the matrix of scale intercorrelations. Four factors were found which, after rotation to orthogonal simple structure, were clearly identifiable as Emotionality (Em), Extraversion-Introversion (EI), Falsification (F), and Sex (Sx). Separate factors of Anxiety and Neuroticism again did not appear in the analysis, and the separation of Emotionality and Extraversion-Introversion as independent factors was confirmed.

References

Bendig, A. W. Extraversion, neuroticism, and manifest anxiety. *J. consult. Psychol.*, 1957, **21**, 398.

Bendig, A. W. Identification of item factor patterns within the Manifest Anxiety Scale. *J. consult. Psychol.*, 1958, **22**, 158.

Bendig, A. W. Score reliability of dichotomous and trichotomous item responses on the Maudsley Personality Inventory. *J. consult. Psychol.*, 1959, **23**, 181. (a)

Bendig, A. W. "Social desirability" and "anxiety" variables in Cattell's IPAT Anxiety scale. *J. consult. Psychol.*, 1959, **23**, 377. (b)

Cattell, R. B. *The neurotic personality factor questionnaire*. Champaign, Ill.: Institute for Personality and Ability Testing, 1955.

Cattell, R. B. A shortened "basic English" version (Form C) of the 16 P-F questionnaire. *J. soc. Psychol.*, 1956, **44**, 257–278.

Cattell, R. B. The conceptual and test distinction of neuroticism and anxiety. *J. clin. Psychol.*, 1957, **13**, 221–233. (a)

Cattell, R. B. *Handbook for the IPAT Anxiety scale*. Champaign, Ill.: Institute for Personality and Ability Testing, 1957. (b)

Edwards, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

Eysenck, H. J. The questionnaire measurement of neuroticism and extraversion. *Riv. Psicol.*, 1956, **50**, 113–140. (a)

Eysenck, H. J. Reminiscence, drive, and personality theory. *J. abnorm. soc. Psychol.*, 1956, **53**, 328–333. (b)

Eysenck, H. G. *The dynamics of anxiety and hysteria*. London: Praeger, 1957.

Frey, A. H., & Becker, W. C. Some personality correlates of subjects who fail to appear for experimental appointments. *J. consult. Psychol.*, 1958, **22**, 164.

Fruchter, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.

Hathaway, S. R., & McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. (Rev. ed.) New York: Psychological Corp., 1951.

Holtzman, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 853–854.

Jensen, A. R. The Maudsley Personality Inventory. *Acta psychol., Hague*, 1958, **14**, 314–325.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187–200.

McCreary, Joyce, & Bendig, A. W. Comparison of two forms of the Manifest Anxiety Scale. *J. consult. Psychol.*, 1954, **18**, 206.

O'Connor, J. P., Lorr, M., & Stafford, J. W. Some patterns of manifest anxiety. *J. clin. Psychol.*, 1956, **12**, 160–163.

Spence, K. W. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Windle, C. The relationships among five MMPI "anxiety" indices. *J. consult. Psychol.*, 1955, **19**, 61–63.

Winne, J. F. A scale of neuroticism: An adaptation of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1951, **7**, 117–122.

(Received April 20, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

PERCEPTION OF MATERNAL CHILD REARING ATTITUDES IN SCHIZOPHRENICS

ALFRED B. HEILBRUN, JR.
State University of Iowa

Theories regarding the conditions antecedent to the development of schizophrenia still remain speculative in character. The points of view fall mainly into three categories: causative factors are of a genetic or constitutional nature; causative factors are of a psychological character linked to some maladjustment in interpersonal relationships; causative factors represent a combination of physical and psychological agents. Many of those who have attempted to account for schizophrenia in terms of psychological factors have stressed the importance of faulty mother–child relationships in the etiology of this disorder (Arieti, 1955; Fromm-Reichmann, 1948; Rosen, 1947). One line of research bearing upon the mother–child hypothesis has utilized an objective measure of the child rearing attitudes of the schizophrenics' mothers, comparing them to the mothers of normals.

Mark (1953) reported that 67 of the 139 items in an attitude survey reliably discriminated between the mothers of male schizophrenics and the mothers of normal males with the schizophrenics' mothers tending to be more "restrictive" and either "excessively devoted or detached." Using Shoben's (1949) Parent–Child Attitude Survey, Freeman and Grayson (1955) found significant or suggestive differences on two out of three subscales when the mothers of schizophrenics and normals were compared. Tendencies to be "possessive" or "ignoring" in attitude were shown by the mothers of schizophrenics. However, a study by Zuckerman, Oltean, and Monashkin (1958a), which compared the responses of these two types of mothers on the Parent Attitude Research Instrument (Schaefer & Bell, 1958), found only one significant difference between them on 20 scales, a finding which can be readily attributed to chance.

The apparent logic behind these studies and numerous others (see Freeman & Grayson, 1955) in which a variety of assessment techniques have been used to evaluate the maternal attitudes of schizophrenics' mothers is that the best way to detect and clarify possible malignant factors in the mother-schizophrenic child relationship is to directly evaluate the mother. However, it can be argued that it is the child's *perceptions* of the mother's characteristics and attitudes which are the important cues mediating his behavior in the mother–child relationship. It would follow that the study of the possible genesis of schizophrenia in disruptive mother–child interactions should proceed by way of the schizophrenic child's perceptual schema of the mother and of himself in relation to her. An earlier study (Singer, 1954) followed a similar approach by using Thematic Apperception Test stories to evaluate schizophrenics' perception of parental characteristics and intrafamilial relationships. Singer concluded that schizophrenics see parental figures as cold or rejecting and depict the relationships with parents as being relatively isolated.

The present study represents an attempt to measure objectively schizophrenic perceptions of mothers' child rearing attitudes and to compare these perceived attitudes with those of normals. More specifically, the hypothesis that schizophrenics would perceive the mother–child relationship as poorer than would normals was tested by predicting that schizophrenics would attribute more pathogenic attitudes to their mothers than would normals.

METHOD

Subjects

Twenty-six females, admitted in 1957 to a typical large midwestern mental hospital and who carried un-

contested staff diagnoses of schizophrenia, were individually administered the Parent Attitude Research Instrument (PARI). There was no attempt made to provide more homogeneous classes of schizophrenics by further categorizing them into the standard subgroups of paranoid, catatonic, etc. One reason for this was the restricted number of schizophrenics available for the study since inclusion demanded the ability and cooperation of *both* the schizophrenic and her mother. The second reason was the lack of any evidence that membership in any subclass was differentially related to the experimental task. A control group of 27 normal females (i.e., answered "No" when asked if they had ever utilized any mental health service) were personally contacted or contacted by mail and were willing to complete the PARI.

The schizophrenic and normal groups did not differ significantly in mean age (schizophrenics, 36.6 years; normals, 33.4 years) or mean educational level (schizophrenics, 11.3 years; normals, 12.1 years). Hospitalization was only partially controlled (26 out of 27 of the schizophrenics were hospitalized at the time of testing while 10 of 26 normals were patients on medical wards of a general hospital) but there is no a priori basis for assuming a *systematic* relationship between the hospitalization factor and the test performance obtained in the present study. The social-economic status variable was not controlled in this study except to the extent that having fairly equivalent mean educational levels for the daughter groups accomplishes this. Since social-economic status has been shown to be related to child rearing attitudes (though not to the *perception* of such attitudes), it seemed necessary to make a preliminary evaluation of whether the mothers of the two daughter groups differed attitudinally. The mothers were also given the PARI and an analysis of variance demonstrated no differences ($F < 1.00$) in expressed child rearing attitudes between the mothers of the schizophrenic and normal daughters used in this study.

Test

The attitude survey used in the present investigation was the Parent Attitude Research Instrument (PARI) developed by Schaefer and Bell (1958). It includes 115 statements devised to measure maternal attitudes judged to be related to personality development in children. The total of 115 statements consists of 23 sets of 5 statements each, with each set intended to measure a category of maternal attitudes. Descriptive scale names are given in Table 1. The test taker has the option of endorsing "strongly agree," "mildly agree," "mildly disagree," or "strongly disagree" for each statement. A "strongly agree" response is given the highest score on the PARI and "strongly disagree" the lowest; therefore, a high score always represents agreement with the attitude in question. Further, the attitudes being measured by 20 of the scales taken in their extreme are typically assumed to be deviant when characteristic of the mother. Accordingly, higher scores for these 20 scales indicate both greater agreement and a greater potential disruptive influence on the child's development. The remaining three "rapport" scales (Nos. 1, 14, 21 on Table 1) were included by the test constructors so that normals would have a greater opportunity to agree with statements on the questionnaire. Scores on these three scales are interpreted in opposite fashion so that higher scores indicate more adjustive attitudes.

Procedure

The schizophrenic daughters were individually administered the PARI under special instructions to "Answer it as you think your mother would." The administration was conducted by some member of the state hospital staff who was previously known to the patient. The staff member was requested to be sure the nature of the task was initially clear to the patient and to restate the directions periodically if it seemed necessary. The 26 schizophrenics included in this study represent about 29% of the 90 female patients with appropriate diagnoses admitted during the data collection period.

The 10 normal daughters obtained from general medical wards were individually administered the PARI by a psychologist under the same conditions as the schizophrenics. The remaining 17 normal daughters were contacted by mail and were strongly urged to complete the task without prior discussion with their mothers concerning it. The names and addresses of these 17 daughters were obtained from their mothers who were initially tested. As far as it is known, the daughters were not acquainted with one another. The mothers, in turn, were obtained as *Ss* from the Iowa City Employment Agency and local civic groups. The total sample of 27 daughters represents about 42% of the 65 requested to participate in the study.

The printed task instructions were identical for both groups, whether seen personally or contacted by letter. Further, the impersonal research nature of the task was stressed for all daughter *Ss* to decrease test response manipulation associated with public description of their mothers.

Scoring of the predictions proceeded just as though the mother had completed the PARI under standard conditions with a score of 4 assigned to a "strongly agree" prediction for a statement, 3 to a "mildly agree," 2 to a "mildly disagree," and 1 to "strongly disagree." These statement scores were then summed over the five statements within each scale. The scale prediction scores on the PARI define perceived maternal attitudes in this study.

Results

The data presented in Table 1 were examined initially by a Type I analysis of variance (Lindquist, 1953) in which the main effect for diagnosis was a between-*S* effect

TABLE 1

COMPARISONS OF MEAN MOTHERS' ATTITUDE SCALE SCORES AS PREDICTED BY SCHIZOPHRENIC AND NORMAL DAUGHTERS

Scale Number and Description	Primary Factor Loading[a]	Schizophrenic ($N = 26$)		Normal ($N = 27$)		t value
		Mean	*SD*	Mean	*SD*	
2 Fostering Dependency	A	14.10	4.00	11.25	3.30	2.85**
3 Seclusion of Mother	A	14.55	3.05	12.45	2.30	2.80**
4 Breaking the Will	A	13.60	3.10	11.25	3.25	2.76**
5 Martyrdom	A	13.75	3.85	10.80	3.60	2.81**
6 Fear of Harming Baby	A	14.80	3.40	14.00	3.30	.89
10 Excluding Outside Influences	A	13.00	2.05	11.45	2.35	2.58**
11 Deification	A	15.35	3.30	13.40	3.30	2.17*
12 Suppression of Aggression	A	13.50	2.75	11.15	1.90	3.62**
15 Approval of Activity	A	14.65	2.65	13.60	3.00	1.31
16 Avoidance of Communication	A	12.90	3.85	10.60	3.10	2.42*
17 Inconsiderateness of Husband	A	13.85	3.55	12.60	4.00	1.19
18 Suppression of Sex	A	13.25	3.50	10.65	3.60	2.74**
19 Ascendance of Mother	A	14.85	3.50	11.80	2.80	3.59**
20 Intrusiveness	A	12.85	3.65	10.60	3.30	2.37*
22 Acceleration of Development	A	15.35	2.60	12.45	3.85	3.05**
23 Dependency of Mother	A	14.20	2.60	12.85	3.40	1.59
7 Marital Conflict	B	14.80	2.75	15.80	3.05	1.25
9 Irritability	B	13.25	2.30	14.80	3.10	2.07*
13 Rejecting Homemaking Role	B	12.00	2.80	11.60	3.95	.42
1 Encouraging Verbalization	C	16.55	2.30	16.75	2.55	.29
14 Equalitarianism	C	16.50	2.60	16.40	2.75	.13
21 Comradeship and Sharing	C	17.85	1.70	18.35	1.45	1.11
8 Strictness		14.45	3.45	13.75	2.95	.78

[a] Taken from Zuckerman, Ribback, Monashkin, & Norton, 1958.
* Significant beyond .05 level.
** Significant beyond .01 level.

and the main effect for scales was based upon within-*S* comparisons. The *F* value for the diagnosis × scales interaction was significant beyond the .001 level of confidence (mean square for interaction = 1.09; mean square for within treatments = .28; $F = 3.89$ for 22, 1122 *df*). Since the diagnosis × scales interaction effect was significant, no evaluation of the main effects of diagnosis or scales was made. Rather, interest was focused upon the simple effects of diagnosis on each of the individual scales (Lindquist, 1953, p. 211). Table 1 presents the comparisons of predicted attitude scores for the individual PARI scales. Inspection of this table shows that in 13 of the 23 scale comparisons, the schizophrenics' prediction scores differed from those of the normals. Of the 13 comparisons providing significant differences, 12 show the schizophrenic daughters ascribing attitudes to their mothers which are more contrary to usually approved child rearing opinions than are those ascribed to their mothers by normals. The exception was the Irritability Scale where normal daughters predicted that their mothers would score higher than did the schizophrenics.

DISCUSSION

When schizophrenic daughters were asked to predict how they thought their mothers would perform on a survey of child rearing attitudes and these predictions were compared with those provided by normal daughters, marked differences in these perceptions were obtained. The major trend was toward the schizophrenic's perception of her mother as endorsing less approved (and presumably more pathogenic) child rearing attitudes when compared to the normal's perception of her mother. This trend was observed for 12 of the 23 PARI attitude scales. Normal daugh-

ters predicted that their mothers would endorse a more pathogenic attitude on only one scale and on the ten remaining scales no significant differences were obtained.

These findings lend support to the notion that the schizophrenic daughter attributes more pathogenic child rearing attitudes to her mother than does the normal daughter. It may be recalled that the mothers of these daughter groups did not differ significantly in their expressed child rearing attitudes on the same attitude scale used by the daughters, a finding which is consistent with the study of Zuckerman et al. (1958a). However, since this study is concerned with perceptual differences between schizophrenics and normals and not with accuracy of perception, this discrepancy between perception and endorsement is not of immediate concern.

Factor analytic studies with the PARI (Schaefer & Bell, 1955; Zuckerman, Ribback, Monashkin, & Norton, 1958b) have found three stable factors which reportedly account for most of estimated communality variance on this instrument. Factor A has been defined as an "authoritarian-control" cluster; Factor B, "hostility-rejection"; and Factor C, "democratic attitudes." An analysis of the findings in the present study shows that 12 of the 16 scales with highest loadings on Factor A (see Table 1) were characterized by a significantly greater perceived pathology in maternal attitudes on the part of schizophrenics. On the other hand, Factor B included scales where schizophrenic–normal comparisons produced a far different picture. On one of the scales (9), normals predicted a reliably higher score for their mothers; the two others (7, 13) produced no differences between the diagnostic groups. Factor C included three scales none of which produced a difference between diagnostic groups. These findings suggest that it is in the "authoritarian-control" aspects of mother–child relationships where schizophrenics perceive their mothers as deviating more from accepted child rearing practices than do normals. Further the "hostility-rejection" dimension either does not distinguish schizophrenics' and normals' perceptions of their mothers or *normals* perceive their mothers as deviating more from standard practices in this regard. Finally, there is no indication of a difference between perceptions of these diagnostic groups with regard to "democratic attitudes" of the mothers.

Thus when the child rearing attitude patterns of the mothers of schizophrenic and normal daughters are examined through the daughters' perception of them and these perceptions are evaluated in rather molar conceptual units, it begins to appear that a major source of interpersonal difficulty in the mother-schizophrenic daughter relationship evolves from the perceived "authoritarian-control" tendencies of the mother.

It might be contended that the findings of the present study are not appropriate evidence regarding the etiological relationship between faulty mother–child relationships and the development of schizophrenia. The argument would be that the perceived child rearing attitudes of the adult schizophrenic are retrospective in nature and might vary markedly from those held as a child (or at least prior to the onset of schizophrenia). It is true that what appear to be systematic perceptual differences between schizophrenics and normals in this investigation can be speculatively related to schizophrenia in a number of ways of which causal significance is but one. In any case, definitive evidence is not to be found in the present study but rather from studies specifically designed to critically evaluate various alternative hypotheses.

It is also recognized that the results of the present study could be accounted for by differential test-taking habits of the schizophrenic and normal daughters rather than differential perceptual habits as previously discussed. For example, if these groups differed substantially along a social desirability response set continuum, one possibility would be that normals would attempt to depict their mothers in a more favorable light than the schizophrenics. Under such a condition, the direction of between-group differences on the PARI would undoubtedly be consistent with those found in this study. However, there are three reasons why accounting for the results in terms of differential social desirability response set tendencies appears less parsimonious than doing so in terms of perceptual differences:

1. The more impersonal research approach was stressed rather than an interest in any given individual's conception of her mother. This might be expected to lessen the likelihood that desirability of response would play a major role within either daughter group.

2. Previous studies comparing psychiatric and normal *Ss* (Gough, 1950; Klett, 1957; McKinley, Hathaway, & Meehl, 1948) indicate that these groups do not show differential characteristics with regard to desirability of test responses.

3. It would be difficult to account for finding systematic differences in predicted attitudes solely within the "authoritarian-control" cluster by presuming differential social desirability response sets. It would seem necessary to introduce a cumbersome triple interaction between social desirability, attitude cluster, and diagnosis to explain why 75% of the scales included in one cluster provided significant unidirectional differences between schizophrenics and normals because of the social desirability factor while not one scale outside of the cluster even approached that result.

Summary

This study compared the perceptions of their mothers' child rearing attitudes on the part of schizophrenic and normal daughters. The two daughter groups, matched for age, education, and partially for hospitalization, were asked to predict how their mothers would respond to items on the Parent Attitude Research Instrument and these predictions operationally defined perceived maternal attitudes. The mothers of the schizophrenics and normals independently completed the PARI, and there were no differences in endorsed child rearing attitudes obtained between these mother groups.

Analysis of the predicted scale scores revealed an overall tendency for schizophrenic daughters to predict attitude scores for their mothers which reflected more variant child rearing practices than did the predicted attitude scores for the normals. These perceived attitude scale scores were further analyzed in terms of three primary factors obtained in previous factor analytic studies. This analysis provided rather convincing evidence that schizophrenics, more than normals, tend to perceive attitudes relating to the "authoritarian-control" dimension as more abnormal in their mothers. However, perceptions of attitudes relating to "hostility-rejection" or to "democratic attitudes" provided no substantial basis for differentiation of these two classes of mothers.

References

Arieti, S. *Interpretation of schizophrenia.* New York: Brunner, 1955.

Freeman, R. V., & Grayson, H. M. Maternal attitudes in schizophrenia. *J. abnorm. soc. Psychol.*, 1955, **50**, 45–52.

Fromm-Reichmann, Frieda. Notes on development of treatment of schizophrenics by psychoanalytic therapy. *Psychiatry*, 1948, **11**, 263.

Gough, H. G. The *F* minus *K* dissimulation index for the MMPI. *J. consult. Psychol.*, 1950, **14**, 408–413.

Klett, C. J. The social desirability stereotype in a hospital population. *J. consult. Psychol.*, 1957, **21**, 419–421.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Cambridge: Riverside, 1953.

Mark, J. C. The attitudes of the mothers of male schizophrenics toward child behavior. *J. abnorm. soc. Psychol.*, 1953, **48**, 185–189.

McKinley, J. C., Hathaway, S. R., & Meehl, P. E. The MMPI: VI. The *K* Scale. *J. consult. Psychol.*, 1948, **12**, 20–31.

Rosen, J. N. The treatment of schizophrenic psychoses by direct analytic therapy. *Psychiat. Quart.*, 1947, **21**, 3–37.

Schaefer, E. S., & Bell, R. Q. Development of a parental attitude research instrument. *Child Developm.*, 1958, **29**, 339–361.

Schaefer, E. S., & Bell, R. Q. Parental Attitude Research Instrument (PARI). Normative data. Unpublished manuscript. National Institutes of Health, Bethesda, Md., 1955.

Shoben, E. J., Jr. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.*, 1949, **39**, 101–148.

Singer, J. L. Projected familial attitudes as a function of socioeconomic status and psychopathology. *J consult. Psychol.*, 1954, **18**, 99–104.

Zuckerman, M., Oltean, Mary, & Monashkin, I. The parental attitudes of the mothers of schizophrenics. *J. consult. Psychol.*, 1958, **22**, 307–310. (a)

Zuckerman, M., Ribback, Beatrice B., Monashkin, I., & Norton, J. A. Normative data and factor analysis on the Parental Attitude Research Instrument. *J. consult. Psychol.*, 1958, **22**, 165–171. (b)

(Received April 22, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

AMBIGUITY AND DISCRIMINATING POWER IN PERSONALITY INVENTORIES

WILLIAM E. BROEN, JR.
University of California, Los Angeles

Research and test construction procedures designed to improve personality inventories can be most fruitful if they are guided by a clear understanding of the rationale of personality inventories. However, disagreements concerning the beneficial or deleterious influence of personality inventory item ambiguity (Allport, 1937; Anastasi, 1954; Benton, 1935; Eisenberg, 1941; Elias, 1951; Freeman, 1955; Hutt, 1945; Landis & Katz, 1934; Meehl, 1945; Rotter, 1954; Vernon, 1953) indicate that, at least with regard to item ambiguity, this clarification of rationale has not been fully achieved. This paper is an attempt to arrive at some further clarification of personality inventory rationale through a discussion of item ambiguity and its relationships to discriminating power. The research and test construction implications of these relationships will also be discussed.

Two Types of Ambiguity

The classical position concerning the relation of item discriminating power and ambiguity is given by Allport in his listing of freedom from ambiguity as a requisite to the use of personality inventories. He states this position as follows: "The stimulus-situation is assumed to be identical for each subject, and his response is assumed to have a constant significance" (Allport, 1937, p. 449). As can be seen from the two halves of Allport's statement, item ambiguity can be divided into two types: variation in the subject's interpretation of what is meant by an item and variation in the extent to which different types of persons give similar responses to an item. These two kinds of ambiguity will be referred to as "interpretive ambiguity" and "response ambiguity," respectively.

Interpretive Ambiguity

The most widely held point of view concerning interpretive ambiguity is that freedom from interpretive ambiguity is necessary in order to achieve adequate discriminating power. This point of view was first reflected in the theoretical arguments of Allport (1937) and in the writings of earlier researchers who found interpretive ambiguity in inventory items and criticized personality inventories because of this ambiguity (Benton, 1935; Eisenberg 1941; Landis & Katz, 1934). The assumption that personality inventories must be free of interpretive ambiguity is also stated by Hutt (1945), and is seen in Taylor's (1953, p. 289) comments on the development of the Manifest Anxiety Scale where, without study of item discriminating power, the majority of the test items were rewritten "so as to minimize confusion and misinterpretation." This assumption is most recently reflected in current texts which list interpretive ambiguity as one of the primary problems of personality inventories (Anastasi, 1955, p. 558; Freeman, 1955, p. 499; Rotter, 1954, p. 257). Among the recent writers, this position is stressed most by Rotter (1954, p. 257), who lists similarity of subject interpretation of questionnaire items as one of the three basic assumptions of personality inventories.

Meehl (1945) was among the first to question the necessity for interpretively unambiguous personality inventory items. He takes the position that discriminating power is the prime requirement for inventory statements, and if discrimination is obtained through empirical item selection, the question of whether or not the discriminating items are interpre-

tively ambiguous is essentially an irrelevant question.

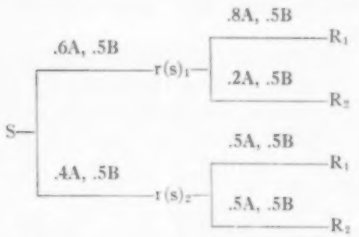
Vernon (1953) also takes the point of view that the question of item ambiguity is not itself a serious question for personality inventories. However, his rationale is different from Meehl's. Vernon (1953, p. 140) sees the basic influence of item ambiguity as harmful but feels that this negative influence is in some way ruled out by the cancelling effects of various interpretive choices over a series of test items. This cancelling effect is based on an assumption that the interpretive choices are randomly distributed among the groups being discriminated.

Elias (1951) goes even further and takes the point of view that interpretive ambiguity is useful and needed for maximal discrimination. His point of view is that the different individual interpretations of inventory statements may themselves reflect relevant personality differences and are thus sources of potentially discriminating variance. "An unstructured question permits greater leeway for personality projection; therefore, it should reflect personality tendencies more sensitively than more structured items" (p. 497).

If the primary task of personality inventories is to classify persons into relevant homogeneous groups, Meehl and Elias are correct in pointing out that amount of interpretive ambiguity is a relevant consideration only to the extent that interpretive ambiguity affects items discriminating power. Criticisms of personality inventories for interpretive ambiguity must then rest on the demonstration of a negative correlation between item interpretive ambiguity and discriminating power.

Partial research evidence on this point is given by Benton (1935), who found no correlation between interpretive ambiguity and discriminating power. These data would seem to contradict the position of those criticizing inventories because of interpretive ambiguity. Benton's data also suggest that there are no interpretive differences between "normal" and "abnormal" subjects, and thus also seem to negate Elias' position that different interpretations yield data relevant to personality categorization. However, the sample of items used in Benton's research is opened to question by Eisenberg's (1941) study where interpretive variations were found to be discriminating. Thus, while suggesting rather clearly that it is not necessary to assume freedom from interpretive ambiguity in order to obtain discrimination, the available research evidence does not go much further in clarifying the relations between discriminating power and interpretive ambiguity. The presentation of a simplified ambiguity model may help in clarifying some of the relations involved.

An ambiguity model. Personality inventory item ambiguity can be most simply represented by a situation in which an inventory item (S) may lead to either of two mediating interpretations ($r(s)_1$, $r(s)_2$), both of which may then be responded to in terms of either Yes (R_1) or No (R_2). As an example, the various probabilities of mediating interpretations and responses for two equal N groups (A and B) being differentiated by an item can be represented as is illustrated below.



The above example assumes that 60% of Group A and 50% of Group B make the first mediating interpretation, and for the As making this interpretation 80% respond Yes (R_1). Multiplying interpretation probabilities by response probabilities we find that in this example 68% of the As and 50% of the Bs answer Yes to the item, and 32% of the As and 50% of the Bs respond No. Thus, if all who respond Yes are labeled as A, the statement correctly identifies 59%.

Trying out various combinations of interpretation and interpretation response probabilities, and studying the resultant discriminating power changes, indicate that the following can be used as a generalized formula for the discriminating power of an item for the model presented above. This formula is appropriate for use with equal or unequal base rate (Meehl & Rosen, 1955) discrimination problems.

$$D = .5 + \left(\frac{|(.5B+.5) - [(PP_1+QQ_1)+B(pp_2+qq_2)]|}{(B+1)} \right)$$

D = proportion correctly identified
B = N for group B/N for Group A
P = Group A probability of $r(s)_1$
p = Group B probability of $r(s)_1$
$Q = 1 - P$
$q = 1 - p$
P_1 = probability of R_1 for the A group $r(s)_1$ persons
p_2 = probability of R_2 for the B group $r(s)_1$ persons
Q_1 = probability of R_1 for the A group $r(s)_2$ persons
q_2 = probability of R_2 for the B group $r(s)_2$ persons

Conclusions concerning interpretive ambiguity and discriminating power. The generalized item discriminating power formula presented above was used to systematically study the influence of base rate changes and changes in interpretation and response probabilities on item discriminating power. The primary conclusions from this study are discussed below.

1. Interpretive ambiguity in itself does not place a limit on item discriminating power. If the mediating interpretations are each perfectly discriminating, the item will be perfectly discriminating regardless of the level of ambiguity. (If $P_1 = Q_1 = p_2 = q_2 = 1.0$; $D = 1.0$ for all values of P and p.) This means that *freedom from interpretive ambiguity is not a necessary assumption of personality inventories,* and research aimed only at describing the extent of interpretive ambiguity in personality inventories is irrelevant to personality inventory utility.

2. If the mediating interpretations are equal in discriminating power, and the majority response is the same for both interpretations, changes in interpretive ambiguity have no effect on item discriminating power. (If $P_1 = Q_1$ and $p_2 = q_2$, D is constant for all values of P and p.) However, in the special case where neither interpretation is highly discriminating and where the majority response is different for the two interpretations (e.g., $P_1 = 1$; $p_2 = 0$; $Q_1 = 0$; $q_2 = 1$) an increase in item ambiguity which leads to different interpretation probabilities for the two groups will lead to increased discriminating power.

3. When the interpretations differ in discrimination power, but the discrimination direction is the same for both interpretations (the direction of the relations between P_1 and $(1 - p_2)$ and Q_1 and $(1 - q_2)$ are the same), then:

(*a*) Increasing item clarity by increasing the probability of the most discriminating interpretation (for the group most discriminated by that interpretation) yields a linear increase in item discriminating power. Table 1 interpretation probability values increasing from .5 through 1.0 give examples of these increases in item discriminating power for several levels of interpretation discriminating power.

(*b*) Increasing item clarity by decreasing the probability of the most discriminating interpretation (for the group most discriminated by that interpretation) yields a linear decrease in discriminating power. Table 1 interpretation probability values decreasing from .5 through .0 provide examples.

These computations suggest an explanation of Elias' data showing ambiguous items as more discriminating than unambiguous items (1951). It is hypothesized that the superiority of Elias' ambiguous items was a function of the increased probability of a more discriminating interpretation rather than a result of the ambiguity itself.

4. The discriminating power of the most differentiating interpretation is the maximum

TABLE 1

CHANGE IN ITEM DISCRIMINATING POWER FOR VARIOUS LEVELS OF INTERPRETATION DISCRIMINATING POWER AND INTERPRETATION PROBABILITY

($p = p_2 = q_2 = .5$; $B = 1$)

Interpretation Probability (P)	Item Discriminating Power[a]		
	$P_1 = .6$	$P_1 = .8$	$P_1 = 1.0$
.0	.50	.50	.50
.2	.51	.53	.55
.4	.52	.56	.60
.5	.525	.575	.625
.6	.53	.5	.65
.8	.54	.62	.70
1.0	.55	.65	.75

[a] If $P_1 = .5$, item discriminating power will equal .5 for all values of P.

potential discriminating power of an item. This maximum discrimination level is approached as item wording changes increase the probability of this most discriminating interpretation. This conclusion suggests a change in the usual inventory construction procedure of empirical selection of statements from an a priori item pool. Increasing discriminating power would be possible by empirical selection among the various statement wordings suggested by interpretations reported by respondents. (The exception to this general rule is the special case under 2 above.)

5. When mediating interpretations differ in discrimination direction, changes in the probability of an interpretation can cause a shift in the discrimination direction of the item. For example, consider the MMPI item "I know who is responsible for most of my troubles." This statement has two common interpretations. One is to include one's self as a causal factor. The other is to interpret the item as referring only to others as causing one's troubles. Let us assume, as seems reasonable, that the first interpretation (including one's self) yields a greater percentage of Yes responses in normals than among paranoids. (Assuming $P_1 = .5$, $p_2 = .2$.) Because of the paranoid's greater tendency to project causation, it is likely that the second interpretation (excluding one's self) yields a greater percentage of Yes responses among paranoids. (Assume $Q_1 = .8$, $q_2 = .5$) MMPI respondent reports suggest that as the item is now worded, the interpretation of excluding one's self as a causal factor has the higher probability. In an equal base rate population a Yes respondent should then be classed as abnormal. (This is what is done in practice. A Yes respondent to this item is given an additional point on the *Pa* scale.) However, if a clinician is disturbed by variance in the interpretations reported by his patients, and "clarifies" the item by changing its wording to "I am responsible for my troubles," because of a belief that "insight" is the variable the test constructor had been trying to get at, the discrimination direction of the item would shift. After this wording change, Yes respondents should be labeled as normal. The relation of discrimination power and direction to interpretive ambiguity for the probability values used in this example is illustrated in Fig. 1.

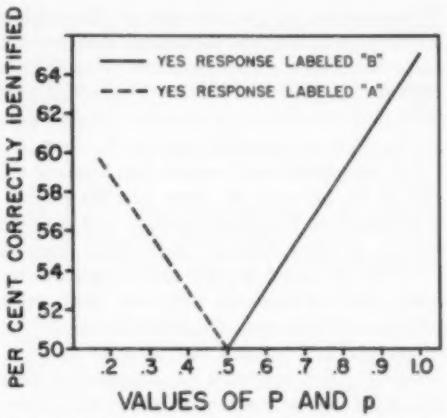


FIG. 1. Shift in discrimination direction occurring as a function of ambiguity lessening changes in item wording. ($B = 1$, $P_1 = .5$, $Q_1 = .8$, $p_2 = .2$, $q_2 = .5$; P and p values of 1.0 indicate an unambiguous item.)

School and other institutional administrators occasionally suggest item wording changes before granting approval to use subjects for research projects. Although the changes suggested usually seem minor, the shifts in discrimination direction, which can be caused even by "clarifying" wording changes, make such changes a very dubious research procedure.

Taylor's MAS wording changes (1953) are another example of clarifying items without adequate study of resultant shifts in item discriminating power and direction. Taylor tested the comparability of the old and revised MAS forms by studying their correlation in a college population. Taylor states the rationale for the use of this population as follows:

> A sample was selected from the college population for this purpose since it was thought that this group would show the least confusion in interpreting the original versions of the difficult items and, therefore, better demonstrate the comparability of the two forms than less verbally sophisticated individuals (Taylor, 1953, p. 288).

Wording changes designed to elicit a greater probability of the high intelligence group interpretation of the original items (judges selected these revised items for nearness to their

interpretation of the meaning of the original item) were thus tested for comparability to the older items on a group explicitly selected for a high probability of the same interpretation of both item forms. Such a method certainly does not provide an adequate test of the comparability of original and revised test forms if the test is to be used in more general population samples. This item revision and revision testing method has not influenced research conclusions based on the MAS because the great bulk of research has been done using only the revised form. However, the same type of revision and revision testing done on a personality inventory with an interpretive base grounded on research with the initial form, would be likely to lead to incorrect conclusions if the old form interpretations were carried over to the revised form. These comments apply to demonstrations of the comparability between any two test forms. *Whenever there is a possibility of interpretive ambiguity, and interpretation choices are not assumed to be random, the comparability of two forms of the same test must be demonstrated on a population with relevant base rates comparable to the base rates in the population in which the tests are to be used.*

6. Use of personality inventories in populations with different base rates can change item discrimination direction and power. For example, at P and p levels of .5 for the item used in Fig. 1, 55% correct identification can be obtained with a $\frac{2}{3}$ A group base rate by labeling Yes respondents as A. However, with a $\frac{2}{3}$ B group base rate, Yes respondents should be labeled as B, and with a .5 base rate for each group, the item is nondiscriminating. These considerations are especially important because of the almost invariable base rate changes between test development populations and populations in which the test is used (Broen, 1957). Also, *unless the interpretation probabilities (P, p) and interpretation response probabilities (P_1, p_2, Q_1, q_2) are known for each group, the effects of known base rate changes on item discrimination direction and power cannot be computed.* Three alternatives are then open to test constructors and validators: (*a*) accept introspective reports of item interpretations as valid self-reports and compute interpretation and interpretation response probabilities from these introspective reports; (*b*) study item discriminating power in populations with relevant base rates comparable to those in the population in which the test is to be used; (*c*) use unambiguous (single interpretation) items. The interpretation response probabilities for the group to be discriminated can, of course, be easily computed for unambiguous items. The utility of using respondent reports as ambiguity measures (in *c*) and interpretation discriminators (in *a*) need not be only an assumption. The utility of the respondent reports can be demonstrated by research dealing with the accuracy of discriminating power predictions using the probability values obtained through the respondent reports for populations with various base rates. Such research would also yield data relevant to Vernon's (1953, p. 140) assumption that the interpretations are distributed randomly throughout various diagnostic classifications. The greater ease of (*c*) over (*a*) and the greater applicability to all populations with known base rates of (*c*) over (*b*) may be the strongest argument for freedom from ambiguity as a criterion in item selection.

Response Ambiguity

The second type of item ambiguity—response ambiguity—has been defined above as the extent to which groups to be discriminated give similar responses and is, therefore, the inverse of item discriminating power. Because of this direct relation to item validity, personality inventory criticisms based on response ambiguity (Allport, 1937; Anastasi, 1954; Eisenberg, 1941; Freeman, 1955; Rotter, 1954) would seem to be more directly relevant to personality inventory utility than criticisms based on interpretive ambiguity. However, methods for combining items of relatively low discriminating power to make highly discriminating scales are available. These methods involve using item configurations (Meehl, 1950); using nondiscriminating items as suppressors of irrelevant variance (Horst, 1941); or combining items which have somewhat higher correlations with the criterion than with each other. An example of the latter type of scale is provided by the

MMPI *Hy* scale which combines two relatively uncorrelated sets of items, each of which provides increased discrimination to the total scale by differentiating on separate dimensions (McKinley & Hathaway, 1944). These methods which enable the use of items of low criterion discriminating power to constrict highly discriminating scales indicate that, like freedom from interpretive ambiguity, freedom from response ambiguity is not a necessary assumption for personality inventories.

SUMMARY

An ambiguity probability model was introduced to clarify the relation of personality inventory interpretive ambiguity to discriminating power. Item discriminating power was found to be a function of interpretation probability, interpretation discriminating power, and population base rates. Research and test construction implications of these findings were discussed, and several methods of using relatively nondiscriminating test items to increase scale discriminating power were cited.

REFERENCES

ALLPORT, G. W. *Personality: A psychological interpretation*. New York: Holt, 1937.

ANASTASI, ANNE. *Psychological testing*. New York: Macmillan, 1954.

BENTON, A. L. The interpretation of questionnaire items in a personality schedule. *Arch. Psychol., N. Y.*, 1935, No. 190.

BROEN, W. E., JR. The use of psychometric sign tests in different populations. *Psychol. Rec.*, 1957, **7**, 117–118.

EISENBERG, P. Individual interpretation of psychoneurotic inventory items. *J. gen. Psychol.*, 1941, **25**, 19–40.

ELIAS, G. Self-evaluation questionnaires as projective measures of personality. *J. consult. Psychol.*, 1951, **15**, 496–500.

FREEMAN, F. S. *Theory and practice of psychological testing*. New York: Holt, 1955.

HORST, P. The prediction of personal adjustment. *Soc. Sci. Res. Coun. Bull.*, 1941, No. 48, 1–156.

HUTT, M. L. The use of projective methods of personality measurement in army medical installations. *J. clin. Psychol.*, 1945, **1**, 134–140.

LANDIS, C., & KATZ, S. E. The validity of certain questions which purport to measure neurotic tendencies. *J. appl. Psychol.*, 1934, **18**, 343–356.

MCKINLEY, J. C., & HATHAWAY, S. R. The MMPI: V. Hysteria, hypomania and psychopathic deviate. *J. appl. Psychol.*, 1944, **28**, 153–174.

MEEHL, P. E. The dynamics of "structured" personality tests. *J. clin. Psychol.*, 1945, **1**, 296–303.

MEEHL, P. E. Configural scoring. *J. consult. Psychol.*, 1950, **14**, 165–171.

MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. *Psychol. Bull.*, 1955, **52**, 194–216.

ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

VERNON, P. E. *Personality tests and assessments*. New York: Holt, 1953.

WELSH, G. S., & DAHLSTROM, W. G. *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.

(Received May 4, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

IMPLICIT PERSONALITY THEORIES OF CLINICIANS AS DEFINED BY SEMANTIC STRUCTURES[1]

MAURICE KORMAN

Southwestern Medical School, University of Texas

The consistently poor showing of clinicians in large scale global assessment studies (Kelly & Fiske, 1951; Holtzman & Sells, 1954) has led to increased concern over the bias introduced by constant processes in the assessor, his reliance upon the connotative meaning of key terms and his idiosyncratic use of concepts.

In his analysis of certain judgmental variables, Cronbach (1955) has shown that if it is a judge's expectation that certain traits covary and others not, he will customarily report subjects to be high (or low) on all of a set of traits. Such a tendency, suggested to reflect an assessor's implicit personality theories, has been demonstrated in different types of judges in diverse settings (Jones, 1954; Soskin, 1954; Steiner, 1954). The psychologist, psychiatrist, or social worker who first observes, tests, or interviews and then applies labels or makes decisions is especially deserving of such attention to his semantic space since many of the concepts he uses are characterized by the looseness and ambiguity of their definitions (Grayson & Tolman, 1950) and their lack of integration in a testable nomological network (Feigl, 1949). Moreover, since the multidisciplinary "team approach" to the patient has found wide acceptance, the lack of semantic clarity and precision has been aggravated by problems of interdisciplinary communication. By virtue of their different training and interests, and because of the partly distinct role they assume, it would seem reasonable to expect psychiatrists, social workers, and psychologists to base their inferences on distinct types of implicit personality theories.

In this study, the clinician's semantic structure is used to reflect that system of interrelationship of meanings that forms one important aspect of his implicit personality theories. This model, based on Osgood and Suci's *D* measure (1952), represents similarities in connotative meaning between concepts as discriminated by an appropriate form of the semantic differential. As applied to groups, this technique has excellent reliability (Jenkins, Russell, & Suci, 1957) and the model which is based upon it can be described as having good construct validity (Lazowick, 1955; Mowrer, 1953; Osgood, 1952).

This investigation will attempt to deal separately with interstructure differences (between disciplines) and intrastructure differences (among concepts). Two questions will be asked: 1. How similar are the implicit personality theories of psychiatrists, psychologists, and social workers, and with respect to what dimensions do they differ? 2. Does an intensive examination of such a network of associations suggest the existence of meaningful covariations or clusterings of concepts?

METHOD

Subjects

The subjects (*Ss*) were 10 psychiatrists, 10 psychiatric social workers, 10 clinical psychologist trainees, 10 PhD psychologists. The first three groups were chosen from among advanced graduate students or psychiatric residents who were near achieving that arbitrary level of training and experience which their

[1] This paper, presented at the meeting of the Southwestern Psychological Association in Topeka, 1959, represents a portion of a dissertation submitted to the University of Minnesota in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is greatly indebted to his advisor, William Schofield, and to others whose assistance aided in the completion of this study.

TABLE 1

CONCEPTS AND SCALES USED IN THIS ANALYSIS

Concepts (listed in the order of their frequency)		Scales (listed by factors)		
Diagnostic	Therapeutic	Evaluation	Potency	Activity
Anxiety Reaction	EST (Electroshock Therapy)	Good–Bad	Hard–Soft	Active–Passive
Depressive Reaction	Psychotherapy	Kind–Cruel	Masculine–Feminine	Excitable–Calm
Schizoid Personality	Supportive Psychotherapy	Successful–Unsuccessful	Strong–Weak	Fast–Slow
Schizophrenic Reaction	Tranquilizing Drugs			
Schizophrenic Reaction, Paranoid Type	Reassurance			
Somatization Reaction	Uncovering Psychotherapy			
Hysteria	Commitment			
Psychopathic Personality	Group Psychotherapy			
Paranoid Personality	Directive Psychotherapy			
Obsessive-Compulsive Reaction				
Passive-Dependent Personality				

respective disciplines demand of them before they are accorded full professional status. The fourth group, composed of senior clinical psychologists who held the PhD degree for an average of 5.80 years, was added in order to further test the sensitivity of the techniques employed and to increase the number of psychologists to a more stable *N* for the investigation of that group's semantic structure.

Concepts

The terms used in this study represent a sample from an empirically derived pool of 95 terms based on an examination of 80 medical folders from four Minneapolis Psychiatric In- and Out-Patient Services. Of the 31 terms or phrases relating to therapy which were thus recorded, 9 were selected from among the 15 most frequently used ones by discarding 4 somewhat ambiguous (e.g., Active Psychotherapy) or less widely used concepts (e.g., Regressive EST). The 64 diagnostic terms in this pool were likewise ranked in the order of their total frequency of occurrence. The 10 most frequently used diagnostic terms were included in this study, together with Passive-Dependent Personality which was used in spite of its lower frequency count because of its contrast to the other terms. The concepts and scales used in this investigation are reported in Table 1.

Scales

Of the many scales available from the various factorial studies, none have been found to be representative of one factor only, and it has been suggested that the factor loadings of the scales vary with the concepts being rated (Osgood, 1952). Since the *Ss* were busy professionals, it was decided to restrict the number of scales to nine to be chosen from the pool of 20 scales that had been set up for the Minnesota *Atlas* study (Jenkins, Russell, & Suci, 1957) and which are believed to be representative of the major dimensions of meaning. A pilot study determined which of these scales best differentiated this type of concept from one another. The 9 scales chosen for this study were, with one exception, the scales having the largest variance. Although Colorful–Colorless was a more widely used scale than Masculine–Feminine, it was decided to include the latter instead in order to give the three major factors, evaluation, potency, and activity, equal representation in terms of number of scales per factor.

Procedure

After the order of the concepts and of the scales had been randomized, and the poles of five of the nine scales reversed so that scales with the same factor loadings were systematically counterbalanced, the

40 *Ss* were instructed to rate each concept on the nine scales using the entire continuum in reporting their first impression of each term.

Results

Intergroup Differences

Each group's semantic structure was determined as follows. Each group's mean scale profile for each concept was determined and as a measure of the similarity in connotation of meaning of any pair of concepts, Osgood and Suci's *D* was obtained. *D* is defined as the square root of the sum of the square differences between scores on each of the scales for any two concepts. Each group's semantic structure is represented by the matrix of *D*s reflecting the distances between the concepts used. Since the use of Pearson product-moment correlations to gauge comparative similarity in semantic structures has been shown to be a sensitive measure by Osgood and Luria (1954) the correlations between each group's set of *D* values and that of every other group were obtained and are reported in Table 2. As can be seen, none of the intercorrelations between group semantic structures falls below .60. All correlations are significantly different from zero ($p < .001$).

In testing the significance of the differences between product moment correlations reported in Table 2, it was found that the semantic structures of the psychology trainees and the PhD psychologists are significantly closer than (*a*) the psychology trainees and the social workers ($p < .01$), (*b*) the psychology trainees and the psychiatrists ($p < .03$) and (*c*) the PhD psychologists and the social workers ($p < .01$). Moreover, the semantic structures of the PhD psychologists and of the psychiatrists are significantly closer to one another than the semantic structures of the PhD psychologists and the social workers ($p < .02$).

TABLE 2

Intercorrelations of Group *D* Matrices

	PhD Psychologists	Psychology Trainees	Psychiatrists
Psychology trainees	.78		
Psychiatrists	.74	.69	
Social workers	.61	.60	.73

TABLE 3

Number of *D*s Differentiating the Four Groups at the .05 Level of Significance[1]

	PhD Psychologists	Psychology Trainees	Psychiatrists
Psychology trainees	1(1,0)		
Psychiatrists	1(0,1)	1(0,1)	
Social workers	7(5,2)	10(5,5)	3(2,1)

[1] The number of *D*s involving diagnostic concepts are given first within the parentheses, followed by the number of therapeutic *D*s.

In an effort to explicate the semantic differences between social workers and the other disciplines, the four groups were contrasted by means of Mann Whitney *U* Tests on each of the 91 *D*s or "links" constituting their semantic structure. As can be seen from Table 3, 23 *D*s significantly differentiated the four groups from one another at the .05 level of significance. Twenty of these differences are between social workers and the other groups; in contrast, the psychologists and psychiatrists differ significantly on only 3 *D*s. Twelve of the 20 *D*s (60%) involve diagnostic categories while the remaining 8 concern therapeutic terms. Since exactly 60% of the pool of 91 *D*s relate to diagnostic terms, it can be seen that social workers differ from the other disciplines as much with regard to diagnostic concepts as they do with reference to therapeutic phrases. A *U* Test applied to the data of Table 3 indicates that the number of significant differences between social workers and the other disciplines is itself significantly higher than the number of such differences among the remaining three groups ($p = .05$). Table 4 reports specific differences between social workers on the one hand and psychologists and psychiatrists on the other. It should be noted that except for Obsessive–Compulsive/Depressive Reaction the social workers find these pairs of con-

TABLE 4

PAIRS OF CONCEPTS DIFFERENTIATING THE GROUPS AT THE .05 LEVEL OF SIGNIFICANCE

Social Workers vs. Other Three Groups[a]		PhD Psychologists vs. Psychology Trainees vs. Psychiatrists	
Diagnostic	Therapeutic	Diagnostic	Therapeutic
Hysteria/Obsessive–Compulsive Reaction	Directive PRx[b]/Tranquilizing Drugs	Hysteria/Somatization reaction	Reassurance/Supportive PRx
Passive–Dependent Personality/Paranoid Personality	Group PRx/Tranquilizing Drugs	Directive PRx/EST	
Obsessive–Compulsive reaction/Psychopathic Personality	Uncov. PRx/Tranquilizing Drugs		
Somatization Reaction/Paranoid Personality	EST/Tranquilizing Drugs		
Hysteria/Somatization Reaction	EST/Supportive PRx		
Hysteria/Psychopathic Personality			
Somatization reaction/Psychopathic Personality			
Obsessive–Compulsive reaction/Depressive Reaction			

[a] Only 13 out of 20 significant differences are reported here since 5 of the *D*s listed differentiated the social workers from more than one other group.
[b] Psychotherapy.

cepts to be *closer* in connotative meaning than do their colleagues.

The Semantic Structure of Psychologists

Since relatively minor differences were expected between the PhD psychologists and the psychology trainees, these groups were combined in order to add more stability to the results. Because the statistical study of relational measures such as *D* is cumbersome and requires many separate tests of significance for even a small number of concepts, only the nine therapeutic terms were used in this phase of the investigation.

The 36 *D*s between all pairs of the nine therapeutic terms for each of the 20 psychologists were computed, resulting in 20 matrices or 36 sets of *D*s each with an *N* of 20. Each concept can be visualized as a point a certain distance away from each of the other terms and the significance of the difference in distance between different pairs of concepts can be calculated by using each concept in turn as a point of reference. In order to answer the kind of question, "Is supportive therapy closer to reassurance than it is to directive psychotherapy," it was necessary to compute 252 Mann Whitney *U* Tests in all. Of these, 83 were significant at the .05 level, 65 being significant at the .02 level.

The presence of such a large number of significant differences in the distance between therapeutic concepts suggests the existence of clusters of concepts. Examination of the sta-

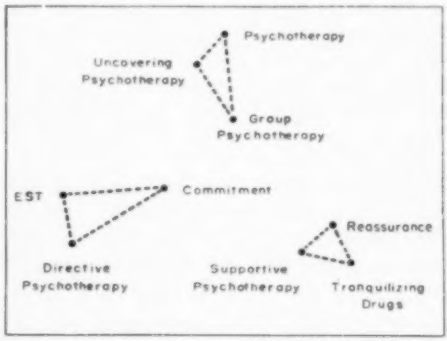


FIG. 1. A two dimensional representation of clusters of therapeutic terms.

tistically significant *D*s between the nine therapeutic terms indicates the existence of three fairly distinct clusters: (*a*) EST, Directive Psychotherapy, and Commitment; (*b*) Uncovering Psychotherapy, Psychotherapy, and Group Psychotherapy; (*c*) Reassurance, Supportive Psychotherapy, and Tranquilizing Drugs. Figure 1 is a two-dimensional visualization of the interrelationships based primarily on topological relations as defined by the tests of significance between *D*s.

Discussion

It is apparent from Tables 2, 3, and 4 that the greatest intergroup differences in implicit personality theories lie between psychologists (regardless of level of experience) and social workers, with psychiatrists maintaining a mediating position between these two disciplines. It must, however, be emphasized that the design of this study did not allow for the testing of the overall importance of these interdisciplinary differences; compared, for instance, to the differences that might possibly emerge from the contrasting of these groups with a group of laymen, such differences as were found in this study might seem minimal and insignificant.

The position of the psychology trainees with respect to the senior psychologist is as expected. There is a communality of interests and outlook between these groups made all the more pervasive by the fact that in this particular sample one group trains and supervises the other. The relationship of both groups of psychologists to the social workers is likewise as hypothesized. Of the three major interests of psychologists, psychodiagnostics, research, and psychotherapy, only the last is shared in some measure by social workers, whose major concern has traditionally been the social welfare of the patient. That psychiatrists should be as "close" to the senior psychologists as to the social workers had not been predicted. One possible explanation would emphasize the fact that the psychiatrist shares both the interest in diagnosis and therapy (as well as in research to an increasing extent) of the psychologist, *and* the social worker's concern with the patient's family, job, financial status, etc. One important inference can further be drawn from these results: Since gross intergroup similarities and differences predicted on the basis of these groups' interests and training are generally reflected in interdisciplinary differences in semantic structures and implicit personality theories, some validity is given to the conclusion that semantic differential ratings and the model upon which the semantic structure is based are sensitive tools and that they parallel extralinguistic interests and attitudes.

In examining the pairs of concepts which social workers find more connotatively similar (Table 4), two alternative explanations offer themselves: (*a*) the two words are actually closer in connotative meaning for social workers or (*b*) social workers do not *discriminate* between the terms, as when all diagnostic terms are uniformly rated "somewhat bad." The second alternative seems the likeliest inasmuch as the pairs of concepts which are connotatively closer for social workers seem to involve constructs that are generally thought to be quite distinct in meaning (e.g., Obsessive-Compulsive reaction/Psychopathic personality). Since these are the only three such differences between PhD psychologists, psychology trainees and psychiatrists, it is difficult to decide whether they represent opposite and interpretable attitudes or chance differences. If these are real differences, the content of the three pairs of concepts (e.g., Hysteria/Somatization reaction) would suggest arguable theoretical differences rather than lack of discrimination. In either case, the semantic structures of psychologists and psychiatrists are seen not to be appreciably dissimilar. The conclusion may, therefore, be drawn that insofar as these semantic structures represent an aspect of the *S*s' implicit personality theories, the systems of covariation of diagnostic and therapeutic terms used by psychiatrists and psychologists are much the same. Social workers, on the other hand, reveal attitudes and associations not shared by psychologists, although accepted to a larger extent by psychiatrists.

Although the intensive analysis performed on the interrelationships among nine terms relating to therapy was based on the semantic structures of psychologists only, it can be seen from the above considerations that the results

could be generalized to the psychiatrists. The first question that presents itself in this attempt to tap some aspects of the clinician's implicit personality theories concerns the basis for the particular clustering that was found. What dimensions account for the intracluster similarities? If one considers the overall median factor scores [2] of terms relating to therapy separately for each of the three factors represented in this study (evaluation, potency, and activity) it becomes clear that Cluster 3 (Reassurance, Supportive Psychotherapy, Tranquilizing Drugs) is "valued" more highly than Cluster 1 (Directive Psychotherapy, Commitment, EST); Cluster 1, however, is more potent or stronger as well as more active. Cluster 2 (Psychotherapy, Uncovering Psychotherapy, Group Psychotherapy) is not as cohesive as the other clusters and it tends to share in the characteristics of the other two groupings.

What attitudes and implicit beliefs relevant to the therapies most commonly prescribed for psychiatric patients can be inferred from this data? It might be argued that the terms in each cluster represent treatments of choice under certain specifiable conditions, other things being equal. Cluster 1, the "strong, active" therapies, might be most applicable to most situations which to the mind of the clinician demanded such rapidly direct and active therapeutic interference at the cost of being forced to sacrifice more highly regarded or valued therapies. Cluster 3, the "good, kind" therapies, would be chosen when a nearly opposite situation existed: the emphasis required would be on a "good, kind, successful" therapy involving relatively little direct interference with the patient's psychopathology. The third cluster, because of its middle position on the three factors, could be said to represent a compromise between more valued therapies and more active and strong ones. Where neither support and reassurance nor active interference are required, Psychotherapy, Uncovering Psychotherapy, and Group Psychotherapy would then seem to be indicated. These terms also differ relatively more among themselves than is the case with the other two clusters. Clinicians seem to attach a valuative connotation to *Psychotherapy* that they find lacking in the other two terms. Is it possible that they have in mind a type of psychotherapy (the one they practice, perchance?) which they prefer to the ones listed here?

The value of such inferences must be tested primarily from their agreement with other knowledge. Insofar as these interpretations clarify the decision making activities of clinicians, and insofar as clinical practice in turn lends credence to them, they contribute to our knowledge of clinicians. Unfortunately, no systematic data is available that reports on clinicians' attitudes towards these therapies. One is therefore forced to fall back on such weak criterion as "do these conclusions make sense in light of how clinicians are commonly believed to feel about these therapies?" The present writer would suggest that extralinguistic attitudes and beliefs concerning therapy are indeed being tapped which are in accordance with the way in which clinicians implicitly match a therapy with their "feeling about the patient." Nevertheless, it cannot be too strongly emphasized that this is an exploratory study in an area in great need of more precise and systematic investigation.

[2] A thesaurus of diagnostic and therapeutic terms based on factor scores is available from the author.

Summary and Conclusions

In an effort to investigate the implicit personality theories of clinicians, 10 clinical psychology trainees, 10 psychiatrists, 10 social workers, and 10 senior clinical psychologists were asked to rate 11 diagnostic and 9 therapeutic frequently used concepts on nine appropriate semantic differential scales. The resulting semantic structures define an aspect of the groups' implicit personality theories. Major conclusions were as follows:

1. Significant intergroup differences in semantic structures and in degree of connotative similarity assigned to pairs of concepts exist between psychologists (regardless of level of experience) and social workers, with psychiatrists maintaining a mediating position between these two disciplines. It is suggested that social workers make fewer discriminations between the connotative meaning of psychiatric and psychologic terms.

2. An intensive analysis of the interrelationships between nine terms relating to therapy for 20 psychologists suggests the existence of three clusters of terms: (*a*) Directive Psychotherapy, Commitment and EST, (*b*) Psychotherapy, Uncovering Psychotherapy and Group Psychotherapy, and (*c*) Reassurance, Supportive Psychotherapy, and Tranquilizing Drugs. These clusters (*a*) can be said to represent the degree of equivalence and interchangeability of the terms in the cluster and (*b*) are suggestive of situations in clinical practice in which the terms of one cluster would be more applicable than those of another. Such clusterings can be presented as providing an insight into the implicit personality theories of clinicians.

REFERENCES

Cronbach, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

Feigl, H. Operationism and scientific method. In H. Feigl & W. Sellers (Eds.), *Readings in philosophical analysis*. New York: Appleton-Century-Crofts, 1949.

Grayson, H. M., & Tolman, Ruth S. A semantic study of concepts of clinical psychologists and psychiatrists. *J. abnorm. soc. Psychol.*, 1950, **45**, 216–231.

Holtzman, W. H., & Sells, S. B. Prediction of flying success by clinical analysis of test protocols. *J. abnorm. soc. Psychol.*, 1954, **49**, 485–490.

Jenkins, J. J., Russell, W. A., & Suci, G. J. An atlas of semantic profiles for 360 words. Univer. of Minnesota Studies on the Role of Language and Behavior, 1957, Tech. Rep. No. 15.

Jones, E. E. Authoritarianism as a determinant of fast impression formation. *J. Pers.*, 1954, **23**, 107–127.

Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. Michigan Press, 1951.

Lazowick, L. On the nature of identification. *J. abnorm. soc. Psychol.*, 1955, **51**, 175–183.

Mowrer, O. H. Changes in verbal behavior during psychotherapy. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research*. New York: Ronald, 1953.

Osgood, C. E. Report on development and application of the semantic differential. Unpublished manuscript, Univer. of Illinois, 1952. (Mimeo.)

Osgood, C. E., & Luria, Z. A blind analysis of a case of multiple personality using the semantic differential. *J. abnorm. soc. Psychol.*, 1954, **49**, 579–591.

Osgood, C. E., & Suci, G. J. A measure of relation determined by both mean differences and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.

Soskin, W. F. Frames of reference in personality assessment. *J. clin. Psychol.*, 1954, **10**, 107–113.

Steiner, I. D. Ethnocentrism and tolerance of trait "inconsistency." *J. abnorm. soc. Psychol.*, 1954, **49**, 349–354.

(Received May 19, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

WISC IQs FOR THE MENTALLY RETARDED

DONALD P. OGDON

Richmond Professional Institute, College of William and Mary

Many psychologists using the Wechsler Intelligence Scale for Children (WISC) have been chagrined to discover that a client has earned an insufficient number of weighted score units to obtain an IQ from the tables provided: the lowest full scale IQ given in the WISC manual is 46, which requires a scaled score of 26 (Wechsler, 1949, p. 26). The primary purpose of this paper is to provide extrapolated IQs down to 28, which is obtained when the scaled score is equal to one.

This extrapolation provides IQs to within the Idiot range, according to Wechsler's (1958, p. 57) classification system. For a variety of reasons, it is considered important by psychologists and others working with the mentally retarded to differentiate among them. Many institutional policies and state legislation require a diagnosis of level of mental retardation or deficiency. This obviously cannot always be provided with the published WISC data. IQs often play a significant role in determining differential therapy, treatment, and training or education in these institutions. However, the principal and most valid reason for desiring to differentiate among individuals with IQs below 46 is to further research and understanding of them. The potential positive contributions to society of these individuals are only beginning to be recognized (Hutt & Gibby, 1958, p. 99).

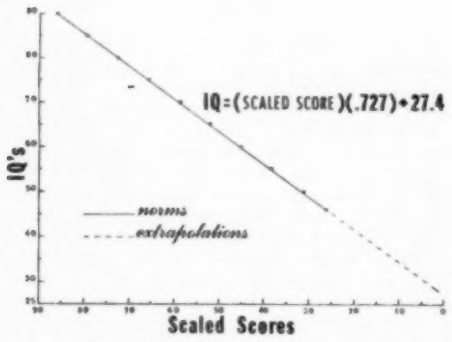


FIG. 1. WISC full scale IQs and scaled scores with extrapolated points below norms.

In Fig. 1, the Full Scale WISC IQs have been plotted against WISC scaled scores. This resulted in a straight line which can be described by the formula IQ = (scaled score) (.727) + 27.4. The formula gives results, when rounded to the nearest whole IQ value, which are identical with those published in the WISC manual for every IQ between 46 and 85 with

TABLE 1

EXTRAPOLATED WISC IQs: IQ = (SCALED SCORE) (.727) + 27.4

Scaled Scores	Calculated IQ Values	Smoothed IQ Values
25	45.58	46
24	44.85	45
23	44.12	44
22	43.39	43
21	42.67	43
20	41.94	42
19	41.21	41
18	40.49	40
17	39.76	40
16	39.03	39
15	38.31	38
14	37.58	38
13	36.85	37
12	36.12	36
11	35.40	35
10	34.67	35
9	33.94	34
8	33.22	33
7	32.49	32
6	31.76	32
5	31.04	31
4	30.31	30
3	29.58	30
2	28.85	29
1	28.13	28

only one exception—with a scaled score of 69, the formula gives the calculated IQ value of 77.56, while Wechsler's Table XI B (1949, p. 26) gives an IQ of 77.

Table 1 presents extrapolated WISC IQs for scaled scores from 1 to 25.

These extrapolated IQ values may be used appropriately only with certain cautions or limitations. Firstly, although the empirical relationship is known to be linear, some relationships that are linear "in the middle range" of their distributions depart from linearity at the extremes.

Generally, the further the extrapolation is made from the empirical data, the greater the probability of error. In this case, as our extrapolated IQs decrease from 46, the less confidence we have in our extrapolation, the lower the expected reliability of the value, and the less its probable psychological significance.

Also attenuating the reliability of the lower extrapolated IQs is the smaller number of items successfully completed to establish the scaled score.

REFERENCES

HUTT, M. L., & GIBBY, R. G. *The mentally retarded child*. Boston: Allyn-Bacon, 1958.

WECHSLER, D. *The measurement and appraisal of adult intelligence*. (4th ed.) Baltimore: Williams & Wilkins, 1958.

WECHSLER, D. *Wechsler Intelligence Scale for Children*. New York: Psychological Corp., 1949.

(Received March 31, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

INCIDENCE OF BENDER-GESTALT FIGURE ROTATIONS

RICHARD M. GRIFFITH AND VIVIAN H. TAYLOR

Veterans Administration Hospital, Lexington, Kentucky

One striking distortion in the drawing of the Bender Visual Motor Gestalt Test (BG) or in the reproduction of block designs—the rotation or angular displacement of one or more of the figures—would be more meaningful to the diagnostician if he had statistics on the frequency of the phenomena in his clinical groups and if he had an adequate theory to elucidate its significance. Aimed at the second problem, a series of attempts to furnish a theoretical underpinning through investigating factors which produce rotations was begun by Shapiro (1953) followed by others (Williams, Lubin, Gieseking, & Rubenstein, 1956; Yates, 1956) who used similar techniques and, most recently, by Hannah (1958) who worked independently. Earlier, Hanvik and Anderson (1950) had considered the effect of focal brain lesions on the production of rotations. The present note provides actuarial statistics on the first problem —the frequency of rotations in the BG drawings of different clinical groups.

METHOD

Drawer by drawer, the files of folders accumulated through eleven years of testing by the Psychology Service of a neuropsychiatric hospital were searched until approximately 1000 BG records (probably one third of all those available!) were examined. The average age of a restricted sample (331 cases) was 35.5 (*SD*, 11.2) at the time the test was administered. The distribution of ages was skewed by the WWI veterans, the median age being 33.4. The IQs of 282 of these patients averaged 94.9 (*SD*, 16.4). As standard operating procedure, looking ahead to future research possibilities, the diagnosis which had been reached by a team composed of psychiatrist, psychologist, and social worker had been noted in each folder. For the present purposes this diagnosis was accepted without further evidence. It is recognized that the design of the study is to some extent impure in that the BG, however small its influence, had been one of the considerations in arriving at the diagnosis.[1]

The diagnoses were broadly categorized into Schizophrenia, Neurosis, Character Disorder, Chronic Brain Syndrome, Mental Deficiency, and Other, the latter being a catch-pile for the less frequent diagnoses and for those instances for which the diagnosis had inadvertently been omitted. When it was noted that the percentage of rotations was high for the first few mental defectives the number of their records was supplemented by combing other drawers of the files; thereafter, their figures were not used in computing the totals for the hospital population. A rotation was defined as an angular displacement of 45 degrees or more in a figure which otherwise remained recognizable.

TABLE 1

PERCENTAGE OF BG RECORDS WITH ONE OR MORE ROTATIONS

	N	Percentage with Rotation(s)
Schizophrenia	346	19.7
Neurosis	167	19.8
Character Disorder	165	16.4
Chronic Brain Syndrome	147	40.8*
Mental Deficiency	59	55.9*
Other	178	23.0
Total[a]	1003	22.8

[a] Total does not include mental defectives.
* Chronic Brain Syndrome and Mental Deficiency differ at .05 probability level; each differs from all other groups at .01 level.

[1] With some clinicians during much of the period covered by the records the BG had been administered primarily as a buffer to what was considered more valid tests. That its effect on diagnoses was never great may be concluded from the breadth of evidence considered by the diagnostic team and from the fact that before they were told the results of the present study a group of 14 psychodiagnosticians within the hospital estimated almost as high per cent rotations by the schizophrenics (36) as by the organics (43).

Results

Table 1 gives the percentage of the records of each diagnostic group in which at least one of the BG designs had been rotated. To the surprise of the investigators and their colleagues, the percentage for the schizophrenics did not differ materially from those for the neurotics and the character disorders (19.7, 19.8, and 16.4, respectively). Above this cluster of values the rotations by the Chronic Brain Syndrome and the Mental Deficiency groups rose to 40.8 and 55.9%. The 23.0% for Others corresponds closely to the 22.8 for the total population (exclusive of the mentally defective).

Discussion

Hanvik and Andersen (1950), requiring less deviation (30 degrees vs. 45) and employing more stringent diagnostic standards, reported a somewhat larger percentage of rotations by organics. Of a group of 44 patients with focal brain lesions which had been verified either through surgery or by Xray the records of 59.1% had one or more rotations as compared with 40.8% for the present group, which was more heterogeneous. Their 18.9% for 37 normals and neurotics is not out of line with the results for the present NP patients—other than organic and deficient.

Because of the relatively small number of mentally defective patients, their percentage rotations is not considered to be as reliable as those for the other diagnoses. Also, in a hospital population the diagnosis is always suspect; mental illness may mimic deficiency. That a relation between intelligence and rotations, as conjectured in the literature (Williams et al., 1956; Yates, 1956), does exist, is supported by comparing the IQs of 56 rotators and 226 nonrotators: the means of 84.3 and 97.5, with standard deviations of 13.3 and 17.1, respectively, differ by the t test at the .01 level.

Summary

This study, confirming the belief of clinicians that rotations of Bender's gestalten are of diagnostic significance, establishes the relative frequency of such distortions by clinical groups within one neuropsychiatric hospital.

REFERENCES

Hannah, L. D. Causative factors in the production of rotations on the Bender-Gestalt designs. *J. consult. Psychol.*, 1958, **22**, 398–399.

Hanvik, L. J., & Andersen, A. L. The effect of focal brain lesions on recall and on the production of rotations in the Bender-Gestalt test. *J. consult. Psychol.*, 1950, **14**, 197–198.

Shapiro, M. B. Experimental studies of a perceptual anomaly: III. The testing of an explanatory theory. *J. ment. Sci.*, 1953, **99**, 394–409.

Williams, H. L., Lubin, A., Gieseking, C., & Rubinstein, I. The relation of brain injury and visual perception to block design rotation. *J. consult. Psychol.*, 1956, **20**, 275–280.

Yates, A. J. The rotation of drawings by brain-damaged patients. *J. abnorm. soc. Psychol.*, 1956, **53**, 178–181.

(Received May 15, 1959)

Journal of Consulting Psychology
Vol. 24, No. 2, 1960

BRIEF REPORTS

A FACTOR ANALYTIC STUDY OF THE SELF-CONCEPT [1]

PHILIP A. SMITH [2]

System Development Corporation, Santa Monica, California

Most measures of the self-concept generate scores which are treated as positive or negative points on a value continuum, despite evidence (Osgood, Suci, & Tannenbaum, 1957) which suggests that a meaningful concept such as the "self" comprises an aggregate of factors rather than a single evaluative dimension. The present study, which was undertaken in an attempt to categorize self-ratings in such a way as to permit a sensitive evaluation of change in psychiatric patients during therapy, casts further light on the factorial structure of the self-concept.

A self-rating device composed of 70 bipolar adjectives descriptive of human personality, including antonyms used by Osgood and terms selected from several popular self-concept measures, was completed by a heterogeneous group of 120 young adult male psychiatric patients. Ratings by each *S* on all 70 scales were dichotomized at the median and a 70×70 matrix of phi coefficients was generated. Two overlapping portions of this matrix were separately factor analyzed and rotated by the quartimax method, an analytic procedure (conveniently programmed for the computer) which provides an objective approximation of orthogonal simple structure. Five interpretable factors were found: Factor I (*self-esteem*) is a broad evaluative dimension characterized at the positive pole by success and satisfaction with life's affairs. It corresponds closely to Osgood's evaluative factors, and appears to be the domain implied by most writers when they refer to the self-concept. Factor II (*anxiety-tension*) is defined by feelings of tiredness, emptiness, fear, and confusion. This dimension, more so than Factor I, may reflect symptoms of distress that are amenable to change during therapy. Factor III (*independence*) appears related to subjectively perceived intelligence and leadership ability. The opposite pole may represent passive acceptance, conformity, and felt inadequacy and inability to lead. Factor IV (*estrangement*) is characterized by feelings of distance and artificiality versus a natural and easy relationship with others. Factor V (*body image*) is similar to Osgood's potency dimension. It is defined by adjectives related to body size and physical strength or potency (e.g., small, light, underweight versus large, strong, rugged, heavy).

Two additional factors were obtained but not interpreted.

The results help explain the findings of investigators who have noted poor correspondence between different tests of the self-concept and low correlation between the self-concept and external criteria of adjustment. It seems likely that these investigators dealt with measures which may have unwittingly confounded several self-concept variables. While it is generally agreed that the extremes of thinking, feeling, and behaving described here are typically observed in psychiatric groups, such states are not *qualitatively* different from those found in normals. Thus, whether "normals" or disturbed *S*s are dealt with, tests which treat self-ratings in simple additive fashion may obscure attributes or dimensions of the self-concept that should be treated independently.

[1] An extended report of this study may be obtained without charge from Philip A. Smith, System Development Corporation, 2500 Colorado Avenue, Santa Monica, California, or for a fee from the American Documentation Institute. Order Document No. 6116, remitting $1.75 for microfilm or $2.50 for photocopies.

[2] Formerly at the Ann Arbor VA Hospital. This paper reports one of a series of studies on the behavior of psychiatric patients carried out by D. W. Bostion, R. Ging, G. L. Hover, and J. J. Lasky of the Psychiatry Research Committee at that hospital. Funds for processing data on electronic digital computing facilities at the University of Michigan Statistical Laboratory were provided by the research committee. The consultation of E. L. Kelly is also gratefully acknowledged.

REFERENCE

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

(Received August 20, 1959)

THE EFFECTS OF CHLORPROMAZINE ON PSYCHOLOGICAL TEST SCORES [1]

ABE J. JUDSON
Utica College of Syracuse University
AND BARBARA W. MacCASLAND
Marcy State Hospital, Marcy, New York

As a consequence of the widespread use of chlorpromazine with new admissions in psychiatric hospitals, patients seen for diagnostic testing have usually had drug medication for several weeks prior to testing. Because the medication obviously introduces a new factor into the testing situation, it is important to evaluate the effect of the drug on test performance. The aim of this investigation is to determine the effects of chlorpromazine on the following tests: the Wechsler-Bellevue, the Rorschach, the Bender-Gestalt, the Porteus Maze, and the MMPI.

Experimental and control groups were selected from new admissions to a state hospital. All *Ss* were administered the test battery within 4 to 10 days after admission and before any drug medication had been prescribed. *Ss* were placed on either chlorpromazine or placebo after completion of the battery and retested 30 days later. Both drug and placebo were discontinued for 48 hours prior to retesting to minimize the influence of the drug's side effects on test performance. There were no significant differences between the two groups in age or education and the modal diagnosis in both groups was schizophrenia. Both male and female patients were included in the sample. There were 18 *Ss* in the drug group, 21 *Ss* in the control group. The chlorpromazine dosage received by each experimental *S* varied with his sensitiviity to the drug but for most *Ss* it was 500 mgs. daily.

The results indicate that under the conditions of the study, chlorpromazine has little or no effect on psychological test scores. On the Wechsler-Bellevue and Bender-Gestalt, there were no significant differences in gains between the two groups. The latter result differs from Winter and Frederickson's (1956) report that the drug impaired Bender-Gestalt performance. However, they attribute their finding to the side effects of the drug and our procedure tended to reduce these effects. On the Porteus Maze, there was some improvement in performance for the control group and virtually no change for the experimental group but the difference in gains falls short of significance ($t = 1.60$). The difference is in the same direction as that reported by Porteus (1957) who found that chlorpromazine reduced Maze Test scores.

The MMPI and Rorschach results are essentially negative. There were no significant differences in gains on any of the MMPI scales. The Rorschach was scored for *W%*, *D%*, *d,dd%*, *F%*, *F − %*, *M%*, *FM%*, *FC%*, *CF,C%*, *A%*, *H%*, *ShR%*, *P*, *R*, and *RT*. Again, there were no significant differences in gains on any of these variables. Although the results are generally negative, it is worth noting that on the *F* scale of the MMPI and *F − %* score of the Rorschach, both measures of unrealistic thinking, the drug group showed greater improvement ($t = 1.83$ and 1.32, respectively). These results are not significant but they do point to a possible effect of chlorpromazine on test performance.

[1] An extended report of this study may be obtained without charge from Abe J. Judson, Utica College of Syracuse University, Utica, N. Y., or for a fee from the American Documentation Institute. Order Document No. 6156, remitting $1.25 for microfilm or $1.25 for photocopies.

REFERENCES

PORTEUS, S. D. Maze test reactions after chlorpromazine. *J. consult. Psychol.*, 1957, **21**, 15–21.
WINTER, W. D., & FREDERICKSON, W. F. The short-term effects of chlorpromazine on psychiatric patients. *J. consult. Psychol.*, 1956, **20**, 431–434.

(Received August 26, 1959)

SIMULTANEOUS RECORDING OF BALES AND CHAPPLE INTERACTION MEASURES DURING INITIAL PSYCHIATRIC INTERVIEWS[1]

A. PAUL HARE, NANCY WAXLER,
Harvard University
GEORGE SASLOW, AND JOSEPH D. MATARAZZO
Medical School, University of Oregon

The content of standardized initial psychiatric interviews is analyzed through the use of the Bales (1950) categories and interaction profile and correlated with simultaneously recorded Chapple measures of the same interviews (Matarazzo, Saslow, & Matarazzo, 1956). Subjects are 24 adults referred to the outpatient psychiatry department in a large hospital. Their average age is 36; 17 are female and 7 male.

The characteristics of the partially standardized interview are described in detail in the article by Matarazzo et al. (1956). The interviewer, an experienced psychiatrist, applies a behavior stress in the second and fourth periods of the interview. The second period consists of a silence stress as the interviewer fails to respond by waiting up to 15 seconds after the subject has finished speaking. In the fourth period the interviewer interrupts the subject for a total of 12 times. Time limits are set on the length of each period.

Two scores which are comparable to the Chapple scores of Units and Action were derived from the Bales data by dividing the total number of acts for a patient into two components. The number of times a patient spoke, his Contributions, was computed by counting the number of times the patient began speaking to the psychiatrist. This score is similar to the Chapple score for Patient's Units. The total number of Bales acts for each patient was divided by the number of Contributions to give a measure of Duration, the average number of acts per contribution. This Duration score based on the Bales data is similar to the Chapple score for Patient Action. No measure similar to the Chapple Silence score appeared to be derivable from the Bales data.

The rank order correlations between the major Chapple variables of Units, Action, and Silence, and three scores derived from the Bales data, Total Acts, Contributions and Duration, are given in Table 1.

The correlation between number of Chapple Patient Units and Bales (Patient) Contributions (+.96) is as high as that obtained for interobserver reliability of Chapple Patient Units for observers using two Interaction Chronographs (Phillips, Matarazzo, Matarazzo, & Saslow, 1957). The correlation between Chapple Patient Action and Bales Patient Duration (+.92) is also high enough to make the two measures interchangeable in the present sample.

[1] An extended report of this study may be obtained without charge from A. Paul Hare, Dept. Social Relations, Harvard University, Cambridge, Mass., or for a fee from the American Documentation Institute. Order Document No. 6154, remitting $1.25 for microfilm, or $1.25 for photocopies.

REFERENCES

BALES, R. F. *Interaction process analysis.* Cambridge: Addison-Wesley, 1950.

MATARAZZO, J. D., SASLOW, G., & MATARAZZO, RUTH G. The interaction chronograph as an instrument for objective measurement of interaction during interviews. *J. Psychol.*, 1956, **41**, 347–367.

PHILLIPS, JEANNE, MATARAZZO, J. D., MATARAZZO, RUTH G., & SASLOW, G. Observer reliability of interaction patterns during interviews. *J. consult. Psychol.*, 1957, **21**, 269–275.

(Received August 29, 1959)

TABLE 1

RANK ORDER CORRELATIONS BETWEEN BALES AND CHAPPLE VARIABLES

($N = 24$)

	2	3	4	5	6
1 Units (Chapple)	−.98	−.14	−.42	+.96	−.93
2 Action (Chapple)		−.05	+.50	−.90	+.92
3 Silence Chapple)			−.35	+.01	+.17
3 Total Acts (Bales)				−.47	+.68
5 Contributions (Bales)					−.91
6 Duration (Bales)					

SOCIAL–PSYCHOLOGICAL FACTORS AND THE PSYCHIATRIC COMPLAINTS OF DISTURBED CHILDREN [1]

RAYMOND G. HUNT,
Washington University
JACK L. ROACH,
New York State Department of Social Welfare
AND ORVILLE GURRSLIN
Youth Board of Buffalo

This study explored relations between symptomatic complaints made by parents about children in outpatient treatment and social status, sex, religion, age, family size, ordinal position, parental age differences, parent–child age differences and family status (presence or absence of both parents).

Data were collected from records of 265 cases in an urban child guidance clinic. Complaints were ordered to seven commonly used categories of disorder: Habit (e.g., enuresis), Personality (e.g., depression), Neurotic (e.g., tics), Conduct (e.g., stealing), Somatic (e.g., asthma), Educational (e.g., poor grades), and Mental deficiency. Frequencies of complaints as a function of the factors under study were analyzed in contingency tables.

Habit disorders were found to be significantly more frequent in younger children and tended to be so among females and among only and oldest children, but were unrelated to any of the other factors. Personality disorders exhibited no relationships to any of the factors under study nor did somatic complaints. Except for a tendency to be more frequent in "normal" as compared with "broken" homes, the same was true of neurotic disorders.

Conduct disorders were significantly more frequent in children from "broken" homes and tended to be so in older children, only children and those with three or more siblings. They were unrelated to the other factors. Educational disabilities were related only to sex, being significantly more frequent in males.

Complaints of mental deficiency were related to most of the factors studied. They were more frequent among intermediate and youngest children and children from "normal" homes. Prevalence was also greater among children farthest in age from their parents and tended to be more frequent among Catholics.

To shed further light on the age relationship here a chi square analysis revealed complaints of mental deficiency to be more frequent among children whose mothers had been 30 years or older at the child's birth. This, of course, is consistent with the well known tendency for the incidence of mental deficiency to increase with maternal age at conception.

Finally, children from "nonnormal" families were significantly overrepresented in the clinic and were significantly older than "normal" family children.

These modest results suggest several possibilities, one of which is that social factors are less important in shaping disorders of children than of older groups. It is also possible though that intake policies and/or cultural attitudes toward psychiatry constitute "filters" producing spuriously homogeneous patient populations.

(Received September 25, 1959)

[1] An extended report of this study may be obtained without charge from R. G. Hunt, Psychology Dept., Washington University, St. Louis 5, Mo. or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 6155, remitting $1.75 for microfilm or $2.50 for photocopies.

Outstanding McGRAW-HILL *Books . . .*

THE NEUROPSYCHOLOGY OF LASHLEY

By FRANK A. BEACH, *University of California, Berkeley*; DONALD O. HEBB, *McGill University*; CLIFFORD T. MORGAN, *University of Wisconsin*; and HENRY NISSEN. Ready in May.

This volume, compiled by four of Lashley's former students, does honor to the memory of Karl S. Lashley—"the man who, as much as anyone, determined the shape of modern psychology." This is a compilation of Lashley's original papers, reflecting the range of his interest and achievements.

CONFLICT, AROUSAL, AND CURIOSITY

By D. E. BERLYNE. *McGraw-Hill Series in Psychology.* Ready in July.

An advanced, graduate-level book intended as a contribution to behavior theory. The author reviews such facts as are available relevant to attention and exploratory behavior; presents a tentative theory to integrate these facts and guide future research; and shows the connections between exploratory behavior and such areas as motivation and learning, thinking, aesthetics, and humor. Further, he points out interrelationships between various lines of theory and research in the field.

BARGAINING AND GROUP DECISION MAKING

By SIDNEY SIEGEL and LAWRENCE E. FOURAKER, both *of Pennsylvania State University.* In Press.

This work was the recipient of the Monograph Prize of the American Academy of Arts and Sciences for the year 1959 in the field of the social sciences. The book is the result of highly original work by a psychologist and an economist into the processes of decision making and choice, including investigation of psychological factors in economic decision. It reports a series of controlled experimental studies on the decisions individuals make when bargaining.

EXPERIMENTS IN HEARING

By GEORG VON BÉKÉSY. *Harvard University. McGraw-Hill Series in Psychology.* 745 pages, $25.00.

An eminent international psychologist and physicist has written this book to make available to physiologists and psychologists the results of his unique ground-breaking experiments in the field of hearing. It is based on most of the papers published by the author during the last 30 years. Half of them were translated from the German.

Send for copies on approval

McGRAW-HILL BOOK CO., INC.

330 West 42nd Street, New York 36, N. Y.

COMPLETE PLAN

AT DEVEREUX SCHOOLS, multidisciplinary teams of rehabilitation-oriented psychiatrists, psychologists, social workers, teachers, and vocational rehabilitation specialists participate in assessment of the whole child and formulate plans for his treatment, education, and vocational training. Children are assigned to one of the many residential units or committees for milieu therapy, general and remedial education, vocational counseling and guidance, pre-vocational services, and vocational training.

CLINICAL STAFF

J. Clifford Scott, M.D.
Edwin H. Abrahamsen, M.D.
Aurelio Buenaanne, M.D.
Charles M. Campbell, Jr., M.D.
Fred J. Calemen, M.D.
Ruth E. Duffy, M.D.
William F. Haines, M.D.
Robert L. Hunt, M.D.
Richard H. Lambert, M.D.
Leonardo Mazzeo, M.D.
Joseph J. Peters, M.D.
Alvin J. Scull, M.D.
Jacob S. Sherman, M.D.
Albert S. Terman, M.D.
Walter M. Uhler, M.D.
Tirso L. Vinuesa, M.D.

Lance Wright, M.D.
V. Ellsworth Henry, S.T.D.
Milton Bratten, Ph.D.
William J. Cohen, Ph.D.
Dorothy E. Conrad, Ph.D.
Sidney L. Cozzi, Ed.D.
Michael S. Dunn, Ph.D.
Shirley M. Johnson, Ph.D.
John R. Kleiser, Ph.D.
Murray Levine, Ph.D.
Henry Platt, Ph.D.
Edgar A. Smith, Ed.D.
George Spivack, Ph.D.
Herbert J. Sprigle, Ph.D.
Anne Rowe, M.S.
Kenneth E. Evans, M.S.

Psychoanalytic Consultants

G. Henry Katz, M. D.
Herbert H. Hornbrook, M.D.

THE DEVEREUX FOUNDATION

A nonprofit organization — *Founded 1912*

Devon, Pennsylvania

Santa Barbara, California — *Victoria, Texas*

SCHOOLS
COMMUNITIES
CAMPS
TRAINING
RESEARCH

HELENA T. DEVEREUX
Administrative Consultant

EDWARD L. FRENCH, Ph.D.
Director

WILLIAM B. LOEB
Treasurer

Professional inquiries for Eastern Schools should be directed to Charles J. Fowler, *Registrar*, Devereux Schools, Devon, Pa.; for Pacific Coast Schools, to Keith A. Seaton, *Registrar*, Devereux Schools in California, Santa Barbara, Calif.; Southwestern residents address Devereux Schools in Texas, Box 856, Victoria, Tex.